Das konzeptuelle Modell des volkstümlichen Märchens und seine maschinelle Umsetzung

# **Einführung**

Diese Arbeit verfolgt das Ziel, einen zuverlässigen Ansatz zur Formalisierung von Märchen zu entwickeln, um deren automatische Analyse zu ermöglichen. Grundlage dieses Ansatzes sind folgende axiomatische Thesen:

**1. ›Handlungsabfolge‹:** Der Inhalt eines Märchens besteht aus einer Abfolge einzelner Handlungen (Aktionen).

**2. ›Handlungstragende Figuren‹:** Jede Handlung im Märchen wird von einer oder mehreren handlungstragenden Figuren geleitet.

**3. ›Auslösung neuer Handlungen‹:** Jeder Auftritt einer Figur oder mehrerer Figuren löst eine neue Handlung aus.

**4. ›Motiv‹:** Das Motiv ist das kleinste untrennbare Inhaltselement im Text und entspricht einer Handlung, die im Verlauf eines Auftritts dargestellt wird. Ein Motiv darf nicht größer als ein Auftritt sein.

**5. ›Markierung des Auftritts‹:** Die Markierung des objektiv erkennbaren Auftrittes im Text entspricht der dem Wechsle zwischen den verschiedenen Handlungsabschnitte und damit den Grenzen zwischen den Motiven.

**6. ›Kategorisierung‹:** Die Kategorisierung der Handlungsabschnitte sowie der beteiligten Figuren definiert inhaltliche Konstanten für das gesamte Genre.

**7. ›Schematisierung der Wechselwirkungen‹:** Die Schematisierung der gesetzmäßigen Wechselwirkungen zwischen diesen Konstanten modelliert eine universelle Struktur des Genres.

Die Feststellung und formale Beschreibung der Inhaltskonstanten und ihrer Wechselwirkungen im Märchen ermöglicht die Entwicklung eines digitalen Assistenten, der imstande ist, entsprechende Inhaltselemente im Märchen zu erkennen und zu annotieren.

Die wichtigste Erkenntnis dieser Arbeit ist die Identifizierung **›**des Auftritts einer handlungstragenden Person**‹** als bisher fehlendes objektives Kriterium zur Erfassung des kleinsten Strukturelements im Märchen, des sogenannten Motivs.

Der erste Teil dieser Arbeit beschreibt das auf diesen Erkenntnissen basierende System der Handlungskategorien und der handlungstragenden Figuren. Anschließend wird versucht, die universelle Struktur des Märchens zu schematisieren.

Im nächsten Abschnitt widmet sich die Arbeit der Bereitstellung von Markup als Mittel zur standardisierten Kodierung des Inhalts. Daraufhin wird das Skript des künstlichen Assistenten vorgestellt, das für die semiautomatische Annotation des volkstümlichen Märchens entwickelt wurde. Zusätzlich wird das Tool zur visualisierten Darstellung der verglichenen Daten präsentiert.

Der Umfang des Begriffs **›**Märchen**‹** wird in dieser Arbeit auf die Bedeutung einer besonderen Gruppe dieses Genres, der sogenannten **›**Zaubermärchen**‹**, reduziert.

# Forschungsstand

Diese Arbeit präsentiert eine digitale Lösung des traditionellen Problems der vergleichenden Märchenforschung. Dabei stehen die Errungenschaften der historisch-geographischen Methode der finnischen Schule sowie die von Wladimir Propp entwickelte strukturalistische Methode im Vordergrund. Die historisch-geographische Methode betrachtet die Struktur des Märchens als eine hierarchische Ordnung bestehend aus den Elementen Motiv, Episode und Typ. Die Terminologie zur Beschreibung dieser Strukturelemente ist oft uneinheitlich, es sei denn, eine einheitliche Gliederungsmethode liegt zugrunde. Die dargestellte Struktur basiert auf dem Systematisierungsversuch von Antti Aarne. Nach seinem Modell setzt sich das Märchen aus folgenden Elementen zusammen:

›**Erzählung**‹**:** Element der obersten Kategorie

**›Hauptteil‹**: Mittleres Element

**›Hauptzug‹**: Kleinstes Element.[[1]](#footnote-1)

Eine weitere Verfeinerung dieser Einteilung geht auf Kaarle Krohn zurück. Er empfahl, die Teile der Erzählung als ›Episoden‹ und die Teile der Episoden als ›Moment‹ zu bezeichnen.[[2]](#footnote-2) Trotz seines Plädoyers, den Begriff ›Moment‹ anstelle von ›Motiv‹ zu verwenden, etablierte sich letztlich der Begriff ›Motiv‹. Schließlich wurden in der vergleichenden Märchenforschung die Begriffe ›Typ‹, ›Episode‹ und ›Motiv‹ zur Unterscheidung der hierarchisch segmentierten Erzählteile verwendet.[[3]](#footnote-3) Das empirische Merkmal, worauf man bei der Definition der deklarierten Systemelemente Bezug nehmen könnte, ist bei keinem Anhänger dieser Systematisierung zu finden.

Der formalistische Ansatz in der Märchenforschung betont die Verallgemeinerung ähnlicher Handlungen im Märchen und spricht von sogenannten Funktionen, die die gesamte Komposition des Märchens bilden. Nach dieser Theorie ist der konkrete Inhalt des Märchens eine Sonderrealisierung der Funktion, die ihrerseits als konstantes Element betrachtet wird. Der Autor dieser Theorie, Vladimir Propp, bietet folgende Beispiele als Anschauungsmaterial an:

»1. Der Zar gibt dem Burschen einen Adler. Dieser bringt den Burschen in ein anderes Reich (171).

2. Der Großvater gibt Sučenko ein Pferd. Das Pferd bringt Sučenko in ein anderes Reich (132).

3. Der Zauberer gibt Ivan ein Kleines Boot. Das Boot bringt Ivan in ein anderes Reich (138).

4. Die Zarentochter gibt Ivan einen Ring. Die Burschen, die in dem Ring stecken, bringen Ivan in das Zarenreich (156). usw.«[[4]](#footnote-4)

Nach den Beobachtungen des Autors sind in diesen Beispielen konstante und variable Größen zu unterscheiden. »Es wechseln die Namen und die entsprechenden Attribute der handelnden Personen, konstant bleiben die Aktionen bzw. Funktionen. Daraus kann man folgern, dass das Märchen häufig völlig gleichartige Handlungen verschiedenen Gestalten zuordnet, wodurch eine Analyse des Märchens auf der Basis der Funktionen der handelnden Personen möglich ist.«

Auch für die Funktion wird im Märchentext kein physisches Merkmal festgelegt, anhand dessen das Vorhandensein dieser Elemente erkannt werden kann. Versucht der Autor später, das Konstrukt für die analytische Segmentierung des Textes zu verwenden, so wird er auf die Notwendigkeit stoßen, neben den Funktionen auch über andere strukturbestimmende Elemente zu sprechen, darunter den Auftritt der handelnden Figur:

»Da wir jetzt wissen, wie die einzelnen Sequenzen angeordnet werden, können wir jedes beliebige Märchen in seine Bestandteile zerlegen. Wir unterstreichen, dass die wesentlichen Bestandteile die Funktionen der handelnden Personen sind. Darüber hinaus gibt es Kopulas und Motivierungen; eine Sonderstellung nehmen verschiedenen Formen des Auftretens der handelnden Gestalten ein (Herbeifliegen des Drachen, Begegnung mit der Hexe), und schließlich gibt es attributive Elemente oder Beifügungen, wie etwa das Hüttchen der Hexe oder ihr Bein als Lehm. Diese fünf Kategorien von Elementen bestimmen nicht nur die Struktur des Märchens, sondern das ganze Märchen überhaupt.«[[5]](#footnote-5)

In Märchentexten gibt es zahlreiche Stellen, bei denen zwei oder mehrere Funktionen in einer Handlung fusionieren. Beispiel: Der Held wird von untreuen Gefährten verraten und in einem tiefen Abgrund zurückgelassen. Dort trifft er ein Männchen, benimmt sich höflich ihm gegenüber und erhält von ihm einen Zauberring, der ihm den Weg in die Oberwelt ermöglicht. Die Handlung der untreuen Gefährten und das daraus resultierende Ergebnis entsprechen der Funktion 203ASchädigung‹ (›A‹). Weitere Funktionen sind: ›Erste Funktion des Schenkers‹ (›Sch‹), ›Reaktion des Helden‹ (›H‹), ›Empfang eines Zaubermittels‹ (›Z‹) und ›Raumverwaltung‹ (›W‹).

Betrachten wir eine andere Variante: Bevor der Held zurückgelassen wird, erklärt ihm die von ihm befreite Frau, wie er den Abgrund mit Hilfe eines Tieres verlassen kann, falls er von seinem untreuen Freund im Stich gelassen wird. Beim Versuch, das richtige Tier zu fangen, steigt der Held auf das falsche und gelangt in eine tiefere Unterwelt.

Auch in dieser Variante tritt die ›Schädigung‹ (›A‹) durch die verräterische Handlung der Gefährten ein. Die Anweisung der geretteten Frau entspricht dem ›Empfang eines Zaubermittels‹ (›Z‹). Eine weitere Handlung, die durch das proppsche Schema nicht erfasst werden kann, endet mit der Landung in der tieferen Unterwelt. Diese könnte zwar als ›Raumverwaltung‹ (›W‹) betrachtet werden, doch muss dabei berücksichtigt werden, dass durch diese Landung eine neue ›Schädigung‹ (›A‹) auftritt. Also im Prinzip handelt es sich um eine Handlung mit zwei Funktionen ›Raumverwaltung‹ (›W‹) und ›Schädigung‹ (›A‹).

Solche Fälle zeigen, dass die von Propp definierten Funktionen nicht den Sequenzen bzw. Handlungsabschnitten entsprechen, sondern deren inhaltlichen Facetten. Was die Handlungsabschnitte angeht, so können sie nicht nur eine, sondern mehrere solcher Facetten enthalten. Davon geht folgendes aus: nicht anhand der Funktionen sind die Handlungsabschnitte zu markieren, sondern umgekehrt, und die Aufgabe der strukturellen Segmentation des Textes besteht darin, die objektiv erkennbaren Marker für den Anfang und das Ende der Handlung im Märchen festzustellen.

Dank der Bestätigung des Grenzübergangs zwischen den Strukturelementen im Text durch regelmäßig auftretende objektive Zeichen unterscheidet sich die in der vorliegenden Arbeit entwickelte Herangehensweise von allen bisherigen Methoden und bietet neue Voraussetzungen für die Diskussion dieses Problems.

# Handlungstragende Figuren im Märchen

Die Formalisierung des Märchens befasst sich mit der Kategorisierung der handelnden Personen und ihrer Handlungen. Dies stellt eine dialektische Einheit zweier Aspekte dar, (a) sozialer Status der handelnden Figuren und (b) ihre funktionelle Rolle im Märchen. Hier im Einzelnen:

## **Sozialer Status der handelnden Figuren**

Bevor wir die handelnden Personen nach ihren Funktionen identifizieren, setzen wir uns mit der gesamten sozialen Ordnung im Märchen auseinander. Basierend auf den Erkenntnissen empirischer Forschung lässt sich die bisher bekannte soziale Ordnung im Zaubermärchen als Zusammenhang folgender Sozialakteure beschreiben:

›Oberhaupt‹: Das Oberhaupt ist in den meisten Fällen ein König oder das Familienoberhaupt; selten tritt an dieser Stelle auch eine Gottheit auf.

›Nachfolger‹: Als Nachfolger gilt das handlungsfähige Kind des Oberhauptes.

›Nachwuchs‹: Der Nachwuchs bezieht sich auf die Vertreter der dritten Generation, zunächst das Kind des Nachfolgers.

›Ebenbürtiger des Nachfolgers, der von außerhalb des Clans bzw. der Familie stammt‹: Der Ebenbürtige ist eine Person mit denselben Rechten wie der Nachfolger, gehört jedoch nicht dem Clan oder der Familie des Nachfolgers an.

›Priester / Magier / Totem‹: Diese Figur zeichnet sich durch ihre besonderen Merkmale, Fähigkeiten oder Fertigkeiten aus. Zum Beispiel ein langer Bart, klein von Wuchs, betagt usw. Ein Magier kann auch der Geist eines verstorbenen Familienmitglieds sein. Oft ist die Angehörigkeit dieser Kategorie durch den Statusbestimmenden Name wie z. B. Fee, Hexe, Zauberer usw. bestimmt. Oft besitzen auch Totems oder personifizierte Naturerscheinungen wie Wind, Frost oder Sonne magische Kräfte.

›Untertan / Berufstätiger / Beamte‹: Typische Untertanen umfassen Vertreter unehrlicher Berufe wie Müller, Hirt, Türmer oder Portier, aber auch andere Berufstätige wie Goldschmied (wobei der Schmied, genauso wie Heilkundiger, eher mit einem Priester verglichen wird), Wirt, Kutscher, Kleinhändler; ebenso wie Beamte, zum Beispiel Minister, Offizier, Berater usw.

›Gegenstand / Gut‹: Dies umfasst materielles und / oder immaterielles Gut, das im Märchen als Objekt der Begierde, als Werkzeug oder als Zaubermittel fungieren kann.

Weiblich vs. Männlich: Jedes Mitglied dieser Ordnung kann durch männliche oder weibliche Wesen repräsentiert werden, was jedoch nicht automatisch für, Gut und Gegenstände gilt, es sei denn, sie sind personifiziert.

## Funktionelle Rolle der handelnden Figuren

Einzelne Vertreter der im Märchen dargestellten sozialen Ordnung können innerhalb der Handlung nur bestimmte funktionelle Rollen einnehmen. Mit anderen Worten: Der soziale Status und die funktionelle Rolle der Protagonisten im Märchen sind eng miteinander verknüpft. Die soziale Rolle gilt als geerbter Status des Protagonisten dessen endgültige Funktion und Status im Märchen durch erworbene Attribute vollbracht werden muss. Für eine detailliertere Diskussion über die funktionellen Rollen der handelnden Figuren ist es wichtig, einen Protagonisten zu identifizieren, der als Maßstab für die Definition der Rollen aller anderen Figuren dienen kann. Als solcher gilt der Held des Märchens, wobei der Begriff geschlechtsneutral verwendet wird.

Das erste wichtige Attribut zur Definition der funktionellen Rolle einer Figur ist der Bezug zur Welt bzw. dem Lager des Helden. Nach dieser Eigenschaft lassen sich die Märchenfiguren in folgende drei Kategorien einordnen:

A. Protagonisten der Welt des Helden (die Unsrigen),

B. Protagonisten der fremden Welt (die Fremden),

C. Weltneutrale Protagonisten (oder diejenigen, die in beiden Welten vorkommen).

Als nächstes Attribut zur Definition der funktionellen Rolle des Protagonisten gilt das Verhalten der Figuren aus den Kategorien (A) und (B) gegenüber den Interessen ihrer jeweiligen Welt. Handelt es sich um eine weltneutrale Figur, wird ihr Verhalten gegenüber den Interessen des Helden beurteilt. Nach dem Wert ihrer Handlung in Bezug auf eigene Welt oder wenn es keine eigene gibt, die Welt des Helden ergeben sich zwei Antipoden für jede Figur: der richtig handelnde Protagonist (kurz r) und das falsch Handelnde Protagonist (kurz f), z. B. der richtige Held und der falsche Held, der richtige Helfer und der falsche Helfer usw.

Tab.1

In der folgenden Matrix versuchen wir, die beschriebene soziale Ordnung und ihre einzelnen Vertreter im Märchen zu präsentieren und zu erklären:

Die obere Hälfte der x-Achse entspricht der Welt des Helden (›Welt der Unsrigen‹).

Der untere Bereich der x-Achse entspricht der fremden Welt.

Rechts von der y-Achse befindet sich das Feld für die richtig handelnden Figuren. Das bedeutet:

* Der Vertreter der ›Welt der Unsrigen‹ handelt zugunsten der eigenen Welt.
* Der Vertreter der ›Welt der Fremden‹ handelt zugunsten der eigenen Welt.
* Die weltneutrale Figur handelt zugunsten der Welt des Helden (der ›Welt der Unsrigen‹).

Links von der y-Achse befindet sich das Feld für die falsch handelnden Figuren. Das bedeutet:

* Der Vertreter der ›Welt der Unsrigen‹ handelt zugunsten der ›Welt der fremden‹.
* Der Vertreter der ›Welt der Fremden‹ handelt zugunsten der eigenen Welt.
* Die weltneutrale Figur handelt zugunsten der ›Welt der Fremden‹.

Die y-Achse zeigt die Protagonisten, die sich falsch oder richtig benehmen können.

Die x-Achse entspricht der Liste der Antipoden jeweils einer von diesen Protagonisten, die die Funktionelle Rolle der im Märchen handelnden Figuren ausmachen. Die Namen dieser Figuren sind jeweils mit zwei Großbuchstaben abgekürzt. Die Kleinbuchstaben ›r‹ bzw. ›f‹ vor den Abkürzungen weisen auf die Attribute ›richtig‹ bzw. ›falsch‹ des entsprechenden Protagonisten hin.

Die Initialen in Großbuchstaben sind wie folgt aufzulösen:

›HH‹: Herr des Helden

›HD‹: Held

›RE‹: Rückeroberungsobjekt

›HF‹: Helfer

›ST‹: Stifter

›VB‹: Verbindende Person

›ZM‹: Zaubermittel

›BZ‹: Besitzer des Zielobjektes

›ZO‹: Zielobjekt

›HP‹: Herr des potenziellen Partners

›PP‹: Potenzieller Partner

›AN‹: Antagonist

Die farbige markierten Zellen in den Feldern und auf der x-Achse zeigen den Bezug zwischen den Mitgliedern der im Märchen vorhandenen Weltordnung und den handlungstragenden Figuren sowie ihre Zugehörigkeit zu eigenen, fremden oder neutralen Welten.

|  |  |
| --- | --- |
|  | Vertreter der ›Welt der Unsrigen‹ |
|  | Vertreter der beiden Welten |
|  | Vertreter der ›fremden Welt‹ |

Im Folgenden betrachten wir diese Figuren anhand ihrer ›richtigen‹ und ›falschen‹ Pendants.

Allgemeine Anmerkung: Es gibt einige Vorbehalte für die gesamte Nomenklatur der handlungstragenden Figuren. Die Berücksichtigung dieser Konventionen ist essenziell notwendig, um sprachliche Unstimmigkeiten zu vermeiden. Handlungstragende Figuren (Personen, Wesen, Gegenstände) stehen überall im Singular und werden durch das grammatische Genus des Terminus im Gegensatz zum Genus und Numerus der Instanz bezeichnet. Zum Beispiel wird für die weibliche Aschenputtel der Terminus ›rHD‹ (richtiger Held) verwendet; die beiden älteren Brüder werden als ›fHD‹ (falscher Held) bezeichnet, d. h. im Singular und im Maskulin, usw.

›HH‹ **–** Der Herr des Helden ist das Oberhaupt seiner Familie. Handelt es sich um eine königliche Familie, so ist der Herr der König. Dieser muss von dem König unterschieden werden, der der Herr des potenziellen Partners (siehe unten) ist und nicht mit dem König gleichzusetzen ist, zu dessen Clan der Held gehört.

Wenn sich der Herr des Helden unverhältnismäßig gegenüber dem Helden verhält, wird er als falscher Herr des Helden (›fHH‹) definiert. Beispiele hierfür sind: Ein Vater begehrt seine eigene Tochter (ATU 706), oder er lässt seine Kinder in der Wildnis zurück (ATU 327A, 707). In all diesen Fällen übernimmt der Familienvater die Rolle eines falschen Herrn des Helden (›fHH‹).

Ein falscher Protagonist ist aus dem instrumentalisierten Protagonisten zu unterscheiden, der seine echte Eigenschaft provisorisch und ausnahmslos nur aufgrund des Druckes von außen auf falsche wechseln muss. Z. B. aufgrund der bösgemeinte Verleumdung seiner älteren Söhne verordnet der König, seinen jüngsten Sohn hinzurichten oder vom Königsreich zu vertreiben (ATU 551).

›HD‹– Wie oben angedeutet, gilt der Held als Maßstab für die Identifikation aller anderen Figuren im Märchen. In der Regel wird der Held durch die Nachfolger vertreten, es sei denn, im Märchen gibt es keinen Herrn des Helden (HH). Im Falle der „gewaltsame“ Kombination von zwei oder mehrere Typen kann ein für einen anderen Status bestimmter Protagonist eine Heldenrolle übernehmen, wie das im angeschlossenen Märchen üblich ist. Folgendes Beispiel: Eine riesige Schlange fordert einem armen Mann, ihm seine Töchter zur Frau zu geben. Die jüngste Tochter stimmt zu der Schlange zu Mann zu nehmen. Nach der Hochzeit stellt die Braut überraschend fest, dass ihr Mann ein prächtiger Jüngling ist, der sich in der Schlangenhaut gesteckt hatte. Bis hierher entwickelt sich die Geschichte so wie es für den Typ ATU 425 – Die Suche nach dem verlorenen Ehemann üblich ist. Weitere Entwicklung dieser Geschichte meint die Verbrennung der Haut und darauffolgende Verschwinden des außergewöhnlichen Bräutigams sowie die von der Frau unternommene Such- und Rettungsaktionen. Statt solcher Entwicklung der Geschichte haben wir in diesem speziellen Fall folgendes zu erfahren: als Zeichen der Dankbarkeit schenkt der Schwiegersohn seinem Schwiegervater einen proviantzeugenden Gegenstand, der ihm ein sorgloses Leben garantieren kann. Die soeben gestartete Geschichte ist als Typ 563 Tischleindeckdich bekannt in der der Held immer derjenige zu betrachten ist, der das Zaubergegenstand zum Geschenk bekommt also, diesmal der Vater und damit der Herr der opferbereitfähigen Tochter also der Heldin wobei dank der besonderen Kombination beide in ungewöhnlicher Rolle dargestellt sind.

Der Held kann eine kämpfende, eine suchende oder auch eine verleumdende und vertriebene Person sein. Nicht selten agieren im Märchen zwei Helden. Als solche gelten: Geschwister (auch Halbbruder oder Halbschwester), ein Ebenbürtiger sowie der Vertreter der dritten Generation, z. B. das Kind eines verdrängten Helden (ATU 707).

Als falsche Helden (›fHD‹) gelten üblicherweise die Geschwister des Helden, sofern sie nicht vor dem Beginn der Heldentätigkeit geboren wurden. Die Rolle des ›fHD‹ kann auch vom Ebenbürtigen übernommen werden, z. B. Schwippschwager aus dem Typ ATU 314.

›PP‹ –Der potenzielle Partner hat denselben sozialen Status wie der Held, stammt jedoch aus der fremden Welt. Gewöhnlich hat der Partner einen eigenen Herrn (HP). Die Beziehung zwischen dem potenziellen Partner (PP) und dem Helden (HD) beginnt immer aufgrund der Initiative einer der beiden Figuren, meistens durch die männliche Figur, selbst wenn die männliche Figur ein potenzieller Partner ist und nicht der Held, wie zum Beispiel im ATU 425. Daher ist es schwierig, ein weiblicher potenzieller Partner von einem weiblichen Zielobjekt (ZO, siehe unten) zu unterscheiden.Letztere kann den Helden zwar heiraten, gilt jedoch nicht als potenzielle Partnerin für den Helden, sondern eher als Beute, die üblicherweise für den Auftraggeber, den falschen Herrn des Helden (›fHH‹) oder den richtigen Antagonisten (›rAN‹) bestimmt ist. Die Initiative, eine solche Figur zu gewinnen, geht vom rAN oder vom fHH aus. Es ist ebenfalls schwierig, einen potenziellen Partner aus dem Rückeroberungsobjekt (›RE‹) zu unterscheiden. Ein Identifikationsmerkmal ist die Heiratstüchtigkeit des Objekts, es muss also nicht der Familie des Helden angehören. Zum Beispiel gilt eine befreite Frau im ATU 301 als potenzielle Partnerin, auch dann, wenn sie zuvor geraubt wurde und der Held den Auftrag des Königs befolgt, um die entführte Frau zurückzugewinnen. Dieses Objekt wird weiterhin als potenzielle Partnerin betrachtet, selbst dann, wenn der Held sie aus verschiedenen Gründen nicht heiraten möchte.

Im Fall eines falschen potenziellen Partners (›fPP‹) handelt es sich in der Regel um die gleichgeschlechtlichen Geschwister (manchmal auch Halbbruder oder Halbschwester) des richtigen potenziellen Partners oder seine Ebenbürtigen. Ein typisches Beispiel wären die ältere Schwester der Königstochter im Märchen ATU 314. Wie bekannt heiraten die ältere Königstochter die Männer aus den vornehmen Familien, wenn die Jüngste zum kahlköpfigen Gärtner wählt der in der Tat ein goldhaariger Jüngling ist der seine Schönheit mit tierischen Magen auf dem Kopf verhüllt.

›HP‹ –Der Herr des potenziellen Partners repräsentiert denselben Stand wie der Herr des Helden, stammt jedoch aus der fremden Welt. Seine Hauptfunktion besteht darin, die Hochzeitsaufgaben und allgemein die Voraussetzungen zu stellen, und wenn nötig, die Freier zu verfolgen. Tut er das nicht oder sorgt sogar dafür das Paar zu unterstützen handelt es sich um den falschen Herr des potenziellen Partners (›fHP‹).

›RE‹ –Das Rückeroberungsobjekt (›RE‹) kann sowohl eine Person als auch ein Gegenstand, ein Tier oder ein Zaubermittel sein. Es handelt sich um das Objekt, das vor dem Beginn der Aktivitäten des Helden in seiner Welt vorhanden war und nun zurückgeholt werden muss. Ist das Objekt nach dem Beginn der Aktivitäten des Helden weggenommen, verschollen oder verschwunden, behält es den Status bei, den es vor der Beginn der Krise hatte, und muss nicht als RE betrachtet werden (beispielsweise bleibt der Schlange-Jüngling aus dem ATU 425 auch dann potenzieller Partner, als er verschwunden ist und von der Gattin (HD) gesucht wird.) Das Vorhandensein des Rückeroberungsbedarfs bereits vor Beginn der Aktivitäten oder der Geburt des Helden oder des Helfers ist wichtig, um das gesuchte Objekt als Rückeroberungsobjekt (›RE‹) zu definieren. Das Rückeroberungsobjekt verhält sich in den meisten Fällen sehr aktiv. Der Held wird beauftragt, die entführte Mutter zurückzuholen. Sie hilft dem Helden dabei, den Aufenthaltsort der externen Seele des Antagonisten auszuspionieren. Das Rückeroberungsobjekt gehört zur eigenen Welt und kann mit verschiedenen Ständen vertreten sein: Nachfolger, beispielsweise die Geschwister, die vom Antagonisten bereits vor der Geburt des Helden festgenommen wurden (›ATU 328‹). Oberhaupt, wenn es sich beispielsweise um die Rückeroberung der entführten Mutter handelt, die jedoch ebenfalls vor dem Beginn der Aktivitäten des Helden entführt sein muss. Jedoch nicht vor der Geburt des Helden, wie das im Typ ›ATU 650A - Der starke Hanns‹ vorkommt. In einigen Fällen ist es schwer, das ›RE‹ von der potenziellen Partnerin zu unterscheiden. Das Mädchen aus dem Typ ›ATU 301‹ kann sowohl Braut als auch geraubtes Objekt sein, das zurückgeholt werden muss. In solchen Fällen ist auf den Bezug zwischen dem betroffenen und anderen Protagonisten zu achten. Wenn ein Protagonist einen eigenen Herrn hat und sich in der fremden Welt befindet, handelt es sich um die potenzielle Partnerin.

Das ›RE‹ kann dann falsch sein, wenn es sich gegen den Helden wendet. Wenn der Held seine, vom ›AN‹ festgehaltenen Brüder befreit und diese ihn dann in den Brunnen werfen, sind die Brüder die Falschen Rückeroberungsobjekte (›fRE‹) und nicht ›fHD‹ (diese können sie sowieso nicht sein, weil sie vor dem Beginn der Aktivitäten des Helden verschollen sind, vgl. ›ATU 303A + 550‹).

›ZO & BZ‹ –Das Zielobjekt und der Besitzer des Zielobjekts. Das Zielobjekt kann sowohl eine Person als auch ein Gegenstand, ein Tier oder ein Zaubermittel sein. Dieses Objekt gehört zu fremder Welt. Das Zielobjekt kann sich sowohl aktiv als auch passiv verhalten. Zum Beispiel der richtige Held (›rHD‹) wird vom falschen Herrn des Helden (›fHH‹) beauftragt, einen Löwen (›ZO‹) zu holen. Der Löwe leistet Widerstand, wird jedoch besiegt und mitgebracht. Der ›rHD‹ wird vom richtigen Herrn des Helden (›rHH‹) beauftragt, ein Pferd (›ZO‹) zu holen. Der Held stiehlt das Pferd, das zwar inaktiv ist, aber dem ›rHD‹ später andere Aufgaben lösen hilft, indem es als Helfer fungiert (vgl. ›ATU 550‹).

In den meisten Fällen haben diese Objekte ihre Besitzer. Der beauftragte Held kommt oft bei der Eroberung des Objekts unmittelbar mit dem Besitzer in Kontakt. Es ist üblich, dass der Besitzer selbst als Zielobjekt angefordert wird, was anschließend zur Bestrafung des Auftraggebers führt (typische Beispiele ›ATU 328, 531‹). Die weibliche Figur des Zielobjekts ist von der potenziellen Partnerin (›PP‹) auch dadurch zu unterscheiden, dass sie keiner Herrschaft untersteht. Nach ihrem Stand soll die weibliche Figur des Zielobjekts eine Magin bzw. ein Gegenstand oder Besitz des Magers sein. Dies bedeutet jedoch nicht, dass alle Figuren im Märchen, die mit magischen Eigenschaften ausgestattet sind, der Kategorie Zielobjekt angehören. Ein Beispiel hierfür ist das Märchen ›ATU 313‹, in dem der Jüngling (›rHD‹/rPP‹) beim Zauberer (›rHP‹/›rHH‹) gerät und sich in dessen Tochter (›rPP‹/›rHD‹) verliebt, mit der er später vor dem Zauberer ›rHP‹/›rHH‹ flieht. Ein Zielobjekt kann auch eine abstrakte Aufgabe sein, wie z. B.: Acker pflügen, Holz holen, Zoll erheben usw., die nicht als Protagonisten des Märchens betrachtet werden.

Ein falsches Zielobjekt liegt vor, wenn das Objekt zugunsten des Helden handelt. Beispielsweise wird nach dem Organ eines Tieres gesucht, und das Tier schenkt dem Helden seinen Jungen. Die Schöne folgt dem Helden, ohne dabei Widerstand zu leisten, usw. Dass diese Objekte später dem Helden helfen, ist jedoch üblich. Im Fall von Zaubergegenständen könnte es sich um ein falsches Objekt handeln, wenn es ausgetauscht wird, zum Beispiel neidische Brüdern wechseln das vom jüngsten Bruder beschaffene Lebenswasser (›rZO‹)durch das salzige Meerwasser aus (›fZO‹).

›ZM‹ –Ein Zaubermittel ist nur dann als solches zu definieren, wenn es nicht als Zielobjekt fungiert. Es muss also nicht gezielt gesucht und erworben werden, sondern kann geschenkt, bestellt, gefunden oder entwendet werden. Der Zauberring, den der Held als Belohnung für seine guten Taten erhält, ist kein Zielobjekt, sondern ein Zaubermittel (ZM). Im Gegensatz dazu ist das Lebenswasser, das der Held im Auftrag seines Vaters beschafft, das Zielobjekt (ZO) und nicht das Zaubermittel, obwohl es tatsächlich magische Eigenschaften hat. Wenn ein echter Ring durch einen falschen ausgetauscht wird, handelt es sich um das falsche Zaubermittel (›fZM‹). Das Gegenteil gilt für den Fall des ausgetauschten Lebenswassers (›fZO‹, siehe oben).

ZM darf man nicht in HF verwechseln. Auch stark personalisierte Zaubermittel die einige Funktionen von helfenden Figuren übernehmen können, bleiben dennoch als ›ZM‹ weil ihr Handeln durch keine Tugend gesteuert wird sondern durch die geerbten Eigenschaften.

›ST‹ & ›HF‹ –Der Stifter und der Helfer. Der Stifter kann sowohl ein Mensch als auch ein Tier, eine Pflanze oder ein mythisches Wesen sein. Der Stifter liefert dem Helden einen notwendigen Gegenstand, eine Fähigkeit oder Informationen, gewährt ihm Asyl bzw. Schütz, mischt sich allerdings nicht in die Handlung des Helden ein. Die von dem Stifter gelieferte Information zeichnet sich dadurch aus, dass sie eine Anweisung, Rat oder Lösungsweg enthält und niemals den Hinweis auf die Schädigung oder Notlage die vom Held beheben werden soll.

Der Helfer (›HF‹) ist jemand, der den Helden begleitet und gemeinsam mit ihm handelt. Das gemeinsame Wirken mit dem Helden bei der Lösung der Aufgabe oder Behebung der Notlage ist ein wichtiger Faktor, um den Helfer vom Stifter zu unterscheiden. Manchmal können die Funktionen von Stifter und Helfer doch in einer Figur verflochten sein, wie zum Beispiel ein Tier-Schwager aus ›ATU 552‹. Er zeigt dem Helden zunächst den Weg zum Antagonisten und übernimmt somit die Funktion des Stifters. Dann jedoch, als der Held im Zweikampf mit dem Antagonisten getötet wird, erweckt ihn derselbe Schwager wieder zum Leben und erfüllt dadurch die Funktion des Helfers.

Nach seinem Stand kann ein Helfer ein Mager oder ein Angehöriger des Magers sein. Stiften und helfen können fast alle Protagonisten im Märchen, einschließlich des Antagonisten. Der Stifter oder Helfer selbst kann jedoch nichts anderes tun als zu stiften und zu helfen.

Der falsche Stifter (›fST‹) bzw. falsche Helfer (›fHF‹) ist jemand, der dem Bösen hilft oder ihm etwas stiftet, zum Beispiel die von einem falschen Helden beauftragte Hexe. Auch die Wesen, die in der fremden Welt beim Besitzer des Zielobjektes wache halten (›ATU 551‹), oder dem Antagonist bei der Verfolgung Hilfe leisten (›ATU 313, 302C\*‹), gelten als falsche Helfer.

›VB‹ –Die verbindende Person wird durch ein menschliches oder mythisches, jedoch eher anthropomorphes Wesen dargestellt. Ihre Hauptfunktion besteht darin, die Vermittlung zwischen dem Helden und einer anderen handlungstragenden Figur zu gewährleisten oder den Helden über die bestehende Krise bzw. notwendige Belastungen zu informieren. Typischerweise werden in der Rolle der verbindenden Person Untertanen eingesetzt, wie etwa eine alte Gastgeberin, ein Wirt oder ein Hirt. Oftmals wird diese Rolle vom Herrn des Helden, meistens von der Mutter übernommen. Also Protagonist mit dem Status ›rHH‹ schließt die Rolle der verbindenden Person (›VB‹) ein.

Eine verbindende Person wird zur Falschen, wenn sie zu Gunsten des Antagonisten handelt. Häufig finden sich in dieser Rolle Diener, Berater usw., die beim bösen König beschäftigt sind. Die Handlung der falschen verbindenden Person ist das Gegenteil dessen, was die Richtige tut (muss aber nicht als solches sein). Als typisch gilt die Rolle eines Denunzianten oder initiierenden der böswilligen Aufträge.

›AN‹ –Antagonist (Gegenspieler, Schädling): Der Antagonist gehört zur fremden Welt und kann praktisch durch alle Stände vertreten werden. Wie oben angedeutet, kann seine Funktion auch von einem falschen Protagonisten, jedoch aus der eigenen Welt, übernommen werden. Besonders verbreitete Fälle sind der falsche Herr des Helden (›fHH‹) und der falsche Held (›fHD‹).

Ein falscher Antagonist, also eine Böse der zu Gunsten des Helden handelt ist nicht anzutreffen aber theoretisch denkbar. In einem Märchen können mehrere Antagonisten auftreten. Wenn der Drache getötet wird und seine Schwester versucht, Rache zu nehmen, gehört auch die Schwester des Drachen zu derselben Kategorie der Handlungstragenden Personen.

Nicht zu übersehen ist einige Überschneidungen zwischen dem vorliegenden Protagonisten-System und der Nomenklatur der handlungstragenden Figuren in der Morphologie des Märchens von Vladimir Propp, sowie in der monographische Abhandlungen der Vertreter der historisch-geographische Methode, allerdings ist dabei folgendes zu betonen: Ausgenommen folgenden vier Figuren HD, HF, ST und fHD ist das vorliegende System völlig anders gestaltet. Das betrifft sowohl die Zahl als auch Definitionen der handelnden Figuren.

# Handlungsabschnitte im Märchen und ihre Organisation

Die Segmentierung des Märchens basiert auf zwei empirisch erkennbaren Entitäten: der Vollgeschichte oder Story und den Motiven. Das eine wird durch den gesamten Text, das andere durch die Auftritte der handlungstragenden Personen im gesamten Text identifiziert.

Im Kontext der Segmentierung von Märchen ist die Rolle der zweiten Entität von entscheidender Bedeutung. Wenn wir berücksichtigen, dass volkstümliche Märchen keine detaillierten Porträts, Naturbeschreibungen oder Gefühlsschilderungen enthalten, wird die universelle Anwendbarkeit des Auftritts handlungstragender Figuren als Maßstab für die Segmentierung von Märchen unbestreitbar.

Auftritt der handlungstragenden Figur im Märchen ist den im Drama explizit markierten Auftritte bzw. Szenen gleichzustellen. Wenn wir ein Märchen als Drama aufzeichnen würden, könnten wir den Text in segmentierte Teile aufteilen, die den Motiven entsprechen. Ein Beispiel dafür ist die segmentierte Zusammenfassung des Märchentyps ›ATU 551 - Das Wasser des Lebens‹ (siehe Abschnitt ›Musterbeispiel‹).

Die Motive bevor sie in die Vollgeschichte integriert werden, lassen sich in dem mittleren Gebilde organisieren. Dieses Gebilde ist traditionell unter dem Namen Episode bekannt.

Durch den Auftritt einer handlungstragenden Figur im Märchen kann eine Episode eingeleitet, fortgesetzt oder abgeschlossen werden. Jede Episode stellt Einheit von folgenden Handlungsphasen dar (siehe Abb.1):

Ein Bild, das Kreis enthält.

Automatisch generierte Beschreibung

Abb.1

Jede Episode im Märchen sowie das ganze Märchen ist mit irgendeiner Krise verbunden. Krise ist ein Zustand, der eine Lösung braucht. Jedes Motiv bezieht sich auf eine Krise und stellt eine Phase ihres Verlaufs dar.

Jedes Motiv ist nach seinem ›Wert‹, ›Sinn‹ und ›Folge‹ zu bestimmen.

Nach dem ›Wert‹ kann ein Motiv positiv, negativ oder ambivalent sein.

Die ›Folge‹ des Motivs bestimmt seine Zugehörigkeit zu bestimmter Episodenphase und somit die gegenseitige Spannung zwischen zwei Motive, die miteinander Konsequenz verbunden sind. Das sind: Offenkundigkeit und Handlung. Die Handlung kann mit zweierlei Ergebnissen dargestellt werden: Lösung oder Auslösung der Krise.

Was den ›Sinn‹ des Motivs angeht, so kann ein Motiv entweder einen Bonus- oder Lastenausgang haben. In einigen Fällen kann ein Ausgang auch beide Sinne haben (darüber unten). Ein Motiv mit Bonus-Ausgang dient der Förderung der handelnden Figur (es stellt ihr Ausrüstung, Gegenstände, Informationen, Versprechen oder Helfer zur Verfügung, die ihr in der Not helfen können). Ein Motiv mit Lastenausgang fordert hingegen den Verbrauch der gesammelten Bonus-Ressourcen, um die Behebung der Krise näher zu bringen.

Der Etalon für die Bestimmung vom ›Wert‹ eines konkreten Motivs ist allein der Faktor des Helden wichtig. Die Beteiligung des Helden an einem Motiv ermöglicht es, den Wert des Motivs entsprechend seiner Auswirkung auf das Interesse des Helden eindeutig zu bestimmen. Problematisch sind hingegen die Motive ohne Teilnahme des Helden. In solchen Fällen hängt die Beurteilung des Motivs ausschließlich von den Ergebnissen der Handlung ab. Hat die Handlung negative Ergebnisse für den Helden, dann wird das Motiv als negativ bewertet. Z. B. verleumden die älteren Brüder den Jüngsten beim Vater so ist das Ergebnis des Motivs eindeutig negativ zu bewerten. Hat das Motiv kein direktes Ergebnis für den Helden, dann wird der Wert des Motivs nach seinem absoluten Wert beurteilt. Z. B. Scheitert ein falscher Held bei der Prüfung des potenziellen Stifters, handelt es sich um ein Bonus-Motiv mit negativem Wert (Grund für diese Bewertung dient die Perspektive des richtigen Helden: für ihn wäre dieselbe Ausgang einem Bonus-Motiv mit einer negativen Folge gleich).

Vor der Offenkundigkeit der ›Krise‹ muss sie (die Krise) erst einmal ausgelöst werden. Die Auslösung der Krise ist das Resultat einer Handlung. Davon geht aus, dass eine Handlung, die eine vorhandene Krise löst, kann gleichzeitig eine neue Krise auslösen oder sie aufdecken (siehe Abb. 1).

Folgendes Beispiel: der vor der schweren Aufgabe stehende Held wendet sich den Stifter an, dem er einmal geholfen hat und bittet ihn um die Gegenleistung. Hiermit handelt es sich um die Offenkundigkeit der Krise für deren Lösung neben dem Helden auch der Stifter mit einbezogen werden muss. Der Stifter händigt dem Helden die Handlungsanweisungen aus, mit deren Hilfe er zur Lösung der vor ihm gestellte Aufgabe näherkommen kann, z. B. den Trick zur Erbeutung eines Reittieres. Der Held folgt der Handlungsanweisung des Stifters und ergreift das Objekt.

Schauen wir uns diese Handlungskette genauer an so wird es uns auffallen, dass das Ausrüsten des Helden mit Handlungsanweisungen nicht nur als Lösung der Krise gilt (Behebung des Mangels an die Erbeutungsmethode), sondern auch als Offenkundigkeit einer bevorstehenden Krise. Für solche Krise ist die Erbeutungsnotwendigkeit des Zielobjektes (hier eines Reittieres) zu halten.

Eine Krise existiert nicht als eigenständiges Motiv, sondern ist entweder die Folge einer Handlung oder die Offenkundigkeit eines solchen Handlungsergebnisses.

Die ›Offenkundigkeit‹ der Krise ist die Situation, in der die Krise bekannt gemacht oder nachvollzogen wird. Die Offenkundigkeit der Krise kann negativ, positiv oder ambivalent sein. Beispiele:

1. Der Zwerg prüft die Höflichkeit der handelnden Figur, um ihr weiterzuhelfen. Hier liegt eine positive Absicht vor, die darauf abzielt, jemandem zu helfen.

2. Ein Mädchen erkennt seinen Makel und wird neidisch auf die Schönheit der Schwester. Dies stellt eine negative Absicht dar, die darauf abzielt, jemanden zu verderben.

3. Auf dem Schild bei einem Kreuzweg steht die Warnung, die ignoriert werden muss, um Erfolg erzielt zu werden. Also der Schild spielt die Funktion einer ambivalenten Offenkundigkeit, das, was richtig ist wird als falsch bezeichnet und umgekehrt.

Der Grund für die Krise und ihre Offenkundigkeit können manchmal ganz verkehrt erscheinen. Ein Beispiel: Der König fühlt sich durch den Helden bedroht (Grund der Krise) und will ihn beseitigen. Deshalb erteilt er dem Helden eine unlösbare Aufgabe, bei deren Erfüllung er sterben soll. Die Offenkundigkeit der Krise lautet jedoch nicht so, wie die Krise heißt, etwa: ›Du musst vom Löwen gefressen werden‹, sondern so: ›Du musst ein Löwenfell besorgen‹. Weil die Offenkundigkeit das Ziel hat, den Helden zu verderben, ist sie trotz ihres Inhalts negativ.

Der Kenntnisnahme der offenkundig gemachte Krise wird durch eine entsprechende ›Handlung‹ gefolgt. Eine positive Handlung schafft entweder einen Bonus für weitere Handlungen oder ermöglicht die Wiedergutmachung der Folgen einer negativen Handlung (Behebung der vorhandenen Krise). Eine negative Handlung lässt hingegen die vorhandene Krise bestehen oder löst eine neue Krise aus (siehe Abb.1). Handlungen mit dilemmatischen Werten können beides tun, wie das Füttern des Vogels mit eigenem Fleisch oder die Tötung eigener Kinder zur Rettung des versteinerten Freundes.

Die Handlung kann verschoben werden. In solchen Fällen wird die Offenkundigkeit einer Krise durch die Offenkundigkeit einer nächsten Krise gefolgt. Es kann auch vorkommen, dass im Rahmen einer Krise begonnene Handlungen durch eine neue Handlung aus dem Rahmen derselben Krise gefolgt werden. Gewöhnlich stellt eine solche Handlung einen ›Mehrwert‹ der vorherigen Handlung dar, insbesondere in sogenannten Bonusepisoden. Obwohl der Mehrwert der Handlung als ein autonomes Motiv etabliert ist, ist er nach seinem ›Wert‹ und ›Sinn‹ von dem vorherigen Motiv nicht zu unterscheiden (daher auch sein Name – ›Mehrwert‹). Anders gesagt, das Resultat dessen, was und wie gehandelt wurde, ist vollkommen von der eigentlichen Handlung abhängig oder sogar in sie beinhaltet. Ein Beispiel: Der ältere Bruder verhält sich frech gegenüber dem Zwerg und verpasst den Bonus von ihm. Das Verhalten des falschen Helden kann hier sowohl die Handlung als auch deren Ergebnis wahrgenommen werden. In einer anderen Variante wird dem falschen Helden nicht nur die potenzielle Hilfe untersagt, sondern er kann gelähmt oder anderweitig verhindert werden (hier gilt die Lähmung als Mehrwert der negativen Handlung).

Die Ergebnisse der Segmentierung des Märchens in Handlungsabschnitte lassen sich wie folgt zusammenfassen:

Die kleinste empirisch erkennbare Entität für die Segmentierung des Märchens entspricht dem Auftritt der handlungstragenden Person(en).

Der Wechsel zwischen den Auftritten in der Erzählung entspricht dem Wechsel zwischen den Segmenten des Märchens. Die kleinste Kategorie dieser Segmente sind Motive, Motive bilden Episoden, die wiederum für die Bildung der Vollgeschichte zuständig sind.

Ein Motiv ist entweder einer Offenkundigkeit oder einer Handlung gleich, dass eine in einer Episode thematisierten Krise widerspiegelt und beginnt und endet im Laufe eines Auftritts.

Ein Auftritt besteht wenigstens aus einem Motiv und kann die Motive von mehr als einer Episode enthalten. Also ein Auftritt ist nicht immer einem Motiv gleich.

Eine Episode setzt sich aus den Motiven zusammen, die die Anfangs- und Lösungsphasen der in ihr thematisierten Krise darstellen.

Motive einer Episode können durch die Motive aus anderen Episoden getrennt werden, daraufhin kann eine Episode entstehen, die praktisch die ganze Geschichte umfasst. Die schematisierte Darstellung einer solchen Konstellation könnte wie folgt aussehen:

[Geschichte [Episode1 [Motiv1.1 + Motiv1.2]] + [Episode2 [Motiv2.1 + Motiv2.2 + Motiv2.3]] + [Episode1[Motiv1.3]]]

Wobei jedes Pluszeichen optional, und jede durch das Pluszeichen getrennte eckige Klammer bindend auf die Grenze zwischen zwei verschiedenen Auftritten hinweist.

# Markup

## Tupel m

Das Markup wird uns helfen, Märchen so zu annotieren, dass alle ihre inhaltlichen Eigenschaften allein anhand dieser Annotationen vergleichbar sind. Die Entwicklung eines solchen Markups ist aufgrund der maximalen Verdichtung der inhaltlich-strukturellen Informationen möglich. Dabei dürfen die Anforderungen an die höhere Ausdruckspotenzial, Anwendungsfreundlichkeit sowie die Kompaktheit des Markups nicht aus den Augen verloren werden.

Zwecks der restlosen Erschließung der von uns oben erfassten Eigenschaften des Motivs (kurz m) versuchen wir, es als ein 4-Tupel von folgenden Attributen darzustellen:

Attribut ›a‹ erfasst die Zugehörigkeit des Motivs zum Typ der Geschichte (ATU–Index)[[6]](#footnote-6), b erfasst die Zugehörigkeit des Motivs zu einer Folge (Handlung oder Offenkundigkeit) sowie zum dessen Wert (negativ, positiv, ambivalent), c fasst den Inhalt des Motivs in einem Wortlaut zusammen, und d erfasst die Nomenklatur der im Motiv beteiligten Figuren.

Hier geben wir die Beschreibung der Variationen jedes dieser Attribute im Einzelnen:

### Attribut a

Jedes Motiv ist Teil einer Episode. Episoden bilden die Vollgeschichte. Also sind jede Episode und damit auch jedes Motiv Bestandteile eines Geschichtentyps. Diese sind, wie bekannt, in einem internationalen Typenkatalog erfasst und entsprechend indiziert.[[7]](#footnote-7) Wir bezeichnen diese Eigenschaft des Motivs mit dem Kleinbuchstaben a (für ATU) und der darauffolgenden Typennummer, zum Beispiel,

›a300 für den Typ ATU 300 – Drachentöter‹.

Dieses Mal beschränken wir die Menge a auf die Vertreter des Subgenres Zaubermärchen. Dieses umfasst die Typen von ATU 300 bis ATU 749.[[8]](#footnote-8)

### Attribut b

Markierung der Zugehörigkeit des Motivs zu dem bestimmten ›Folgen‹ ist durch entsprechende konventionellen Zeichen Möglich:

|  |  |
| --- | --- |
| Zeichenname | Die Folge (Art des Motivs) |
| Ef | Offenkundigkeit der Krise |
| Ha | Handlung in der Krise |

Die Variationen nach dem unterschiedlichen Wert jeder von diesen Folgen lassen sich wie folgt unterscheiden:

|  |  |
| --- | --- |
| Zeichen | Die Folgen (Art des Motivs) |
| F (großes Ef) | Offenkundigkeit der Krise mit positivem Wert |
| f (kleines Ef) | Offenkundigkeit der Krise mit negativem Wert |
| Ff (beide Ef) | Offenkundigkeit der Krise mit ambivalentem Wert |
| H (großes Ha) | Handlung in der Krise mit positivem Wert |
| h (kleines Ha) | Handlung in der Krise mit negativem Wert |
| Hh (beide Ha) | Handlung in der Krise mit ambivalentem Wert |
| HF (großes Ha und kleines Ef) | Handlungsanweisung |
| hF (kleines Ha und großes Ef) | Fehlende oder abzuratende Handlungsanweisung |

Zusammengefasst sieht die Menge der Variationen des Elements ›b‹ wie folgt aus:

Einiges mehr über die Attributwerte ›HF‹ und ›hF‹: Im Gegensatz zu anderen Werten erfassen diese Attribute beide Episodenphasen: ›Ha‹ für Handlung und ›Ef‹ für Offenkundigkeit. Grund dafür ist die oben erwähnten polysemen Natur der Handlungsanweisungen. Diese enthalten sowohl die Informationen, die dem Helden den Lösungsweg erklären oder ihn mit einer Anleitung ausstatten (›H‹), als auch den Auftrag, den der Held erfüllen muss (je nach dem Status des Auftraggebers ›F‹ oder ›f‹).

### Attribut c

Inhalt der Aktion stellt einen Wortlaut dar und zählt somit zum informativsten Bestandteil des Markupelements. Dieser Wortlaut entspricht dem inhaltlichen Wert des im Rahmen einer Episodenphase stattgefundener Aktion, z.B. ›Begehren\_einer\_Person\_oder\_eines\_Gegenstandes‹.

Im Gegensatz zu anderen Attributen des Tupels, deren Variationen bereits vorbestimmt sind, ist eine endgültige Fertigstellung der Wortlautliste, also Variationen für das Attribut ›c‹*,* eher undenkbar. Es kann immer Variationen geben, die neue Formulierung fordern. Bereits aus diesem Grund zählt das Konzipieren eines Modells für die vorschriftliche Herstellung der Wortlautwerte zu wichtigste Aufgabe bei dem Bau des Markups. Ohne Lösung dieser Aufgabe wäre standardmäßiger Einsatz des Markups zum Scheitern verurteilt.

Dank der Nutzung der Tupel-Konstruktion des Markupelements ist es möglich, einen und denselben Wortlaut in verschiedene Fassetten mehr als einmal zu benutzen. Diese Kategorie von Wortlauten nennen wir die ›Klon–Variante‹. ›Klonen‹ oder unveränderte Anwendung eines Wortlautes ist gewöhnlich möglich beim Wechsel zwischen der Offenkundigkeit als Beauftragung und dem Resultat der danach geleisteten Handlung (mit positiven oder negativen Folgen). Die Variation des Tupels ist dabei durch den Wechsel von ›b‹ Attributen zu erzielen. Anhand des Attributes ›d‹ ist außerdem möglich die durch die Klon–Wortlauten erfasste Inhalte weiter zu kontextualisieren.

Folgendes Beispiel: ›F:HEILMITTEL\_besorgen:rHD\_rHH‹.

Das Element erfasst den vom Herrn des Helden (›rHH‹) zum Helden (›rHD‹) mit guter Absicht erteilten Auftrag (›F‹), ihm ein dringend benötigtes Heilmittel zu beschaffen.

Durch den Wechsel von ›F‹ zu ›f‹ und von ›rHH‹ zu ›rAN‹, hätte das Tupel mit derselben Wortlaut einen wesentlich anderen Inhalt zusammenfassen können. Das wäre nämlich der vom Antagonisten mit einer schlechten Absicht erteilte Auftrag, der Held bei der Erfüllung der unlösbaren Aufgabe, hier bei der Beschaffung des Heilmittels, verderben lassen. Das entsprechende Markupelement sieht wie folgt aus: ›f:HEILMITTEL\_besorgen:rHD\_rAN‹.

Durch den weiteren Wechsel der Werten vom Attribut ›b‹ ist außerdem möglich der Mehrwert der positiv, negativ oder ambivalent ausgegangenen Handlung zu bezeichnen. Das folgende Markupelement kann z. B. wie folgt aufgelöst werden: Der Held endet seine Handlung zur Suche des Heilmittels mit positiven Folgen und händigt seinem Beauftragten das benötige Zielobjekt aus:

›H:HEILMITTEL\_besorgen:rHD\_rHH\_rZO‹

Kämme ein Zwischenfall vor, bei dem das richtige Lebenswasser vom falschen Helden (hier von neidischen Brüdern) durch das einfache ausgewechselt wird, so könnte auch diesen Fall durch denselben Wortlaut, mit gewechselten Werten von ›b‹ und ›d‹ Attributen erfasst werden:

›Hh:HEILMITTEL\_besorgen:rHD\_rHH\_fZO‹

Ein anderer Einsatz der Klon–Methode ist für die Zusammenfassung des ersten Treffens zwischen dem Helden und dem Stifter oder dem künftigen Weggefährten möglich. Das Letzte kann sowohl ein Helfer als auch ein falscher Held sein. Bei solchen Fällen handelt es sich gewöhnlich um die Probestellung der handelnden bzw. reisenden Figur und deren Reaktion. Die älteren Brüder antworten frech auf die Frage des ihnen auf dem Weg begegneten Männleins und ernten Misserfolg bei der Erfüllung der Aufgabe. Der jüngste Bruder benimmt sich höflich gegenüber dem Männlein und sichert somit seine Hilfe, die in unterschiedlicher Form auftreten kann. Die Begegnung mit dem Zwerg und deren Vorgang kann man z. B. mit Hilfe folgendes Wortlauts erfassen ›Gutes\_Benehmen‹ Präfixe ›Ef‹ und ›Ha‹ hätten dabei auf die Phase des Treffens hinweisen können: ›F:Gutes\_Benehmen‹ als [verschleierte] Offenkundigkeit der Voraussetzung für die potenzielle Hilfe und ›H: Gutes\_Benehmen‹ als positive Reaktion des Helden auf diese Voraussetzung oder auch ›h:Gutes\_Benehmen‹ als negative, unhöfliche Reaktion des falschen Helden auf derselben Probe.

Nicht alle, miteinander konsekutiv verbundene Phasen einer und derselben Handlung, lassen sich durch das Klonen des Wortlauts erfassen. Um den Zusammenhang zwischen diesen Handlungen bzw. Handlungsphasen auch in diesen Fällen trotzdem fixieren zu können wird sogenannte ›Ableitungsmethode‹ verwendet. Dabei übernimmt ein charakterisierendes Stichwort die Funktion einer Achse, wodurch mehrere, inhaltlich nacheinander folgende Wortlaute trotz ihrer Unterschiede in Verbindung bleiben, (aus technischem Grunde wir eine solche Achse im Wortlaut mit großgeschriebene Buchstaben markiert) z. B.

›h:SCHWEIGEPFLICHT\_verhängen‹ und ›H:SCHWEIGENPFLICHT\_beheben‹

Am häufigsten ist diese Methode bei der Zusammenfassung solcher Handlungsphase einzusetzen wie die Auslösung und die Lösung der Krise (1), oder die Offenkundigkeit einer vorbestimmten Krise und deren Eintritt (2). Beispiele.:

(1) ›h:KONTAKT\_abbrechen‹ , ›F:KONTAKT\_wiederherstellen‹ und ›H:KONTAKT\_wiederherstellen‹

(2) ›h:KRANKHEIT‹, ›F: KRANKHEIT\_Beheben‹ und ›H: KRANKHEIT\_Beheben‹

Stichworte von einigen Ableitungen können mehr als zweimal wiederholt werden z. B. TABU, MISSETAT, VORTÄUSCHEN usw. Es handelt sich um die Serialisierung von einzelnen Stichworten mit dem Ziel die Variationen kompakt zu erfassen. Z. B. Offenkundigkeit der Krise: ›F: TABU\_Behälter\_nicht\_öffnen‹, ›TABU\_Feuer\_darf\_nicht\_erlöschen‹, ›TABU keine Eindringlinge einlassen‹ etc. und die für alle Tabu–Varianten geltende Handlung mit negativer Ergebnis ›h:TABU\_brechen‹.

Neben der Klonierung und Ableitung wird es versucht die Bedeutungen der Prädikate im Auslaut des Wortlauts so zu generalisieren, dass sie so viel wie möglich Fassetten umfassen. Derzeit sind folgende handlungsgeneralisierende Prädikate in Einsatz:

›beheben‹ – gilt als gemeinsames Prädikat für die Offenkundigkeit und / oder Mehrwert der Handlung:

›F: WASSERMANGEL\_beheben‹ und ›H: WASSERMANGEL\_beheben‹

›erhalten‹ – gilt als gemeinsames Stichwort für den Mehrwert der zum Bonusbringenden Interaktion zwischen dem Helden und dem Stifter:

›H:ZAUBERMITTEL\_erhalten‹ und ›H:Flugfähiges\_Transportmittel\_erhalten‹ etc.

›erfassen‹ / ›entgegentreten‹ / ›anpacken‹ – fasst die Handlung zusammen, bei der das Zielobjekt oder sein Besitzer oder die Aufgabe, die kein Objekt hat wird erstmalig erfasst, angepackt oder Entgegentreten (bevor er / sie / es behoben oder besorgt wird):

›HEILMITTEL erfassen‹, ›VERSTECKWETTE anpacken‹, ›Der RESPEKTLOSIGKEIT entgegentreten‹ etc.

›besorgen‹ / ›herbeiholen‹ – bezeichnet einen erfüllten (misslungenen) Auftrag (jedoch nicht das erstmalige Ergreifen des Zielobjektes oder seines Besitzers):

›REITTIER besorgen‹, ›Die SCHÖNE herbeiholen‹ etc.

›nachgehen‹ – gilt als generalisierte Form solcher Verhalten wie: Akzeptieren oder Ablehnen eines Angebotes, sowie eine richtige oder falsche Reaktion auf die Gefahr, Verführung, Provokation oder Vortäuschung:

›h:ungewollter\_VERLOBUNG\_nachgehen‹ und ›H:ungewollter\_VERLOBUNG\_nachgehen‹

›h:Dem\_AMUSEMENT\_nachgehen‹ und ›H:Dem\_AMUSEMENT\_nachgehen‹

›deuten‹ – kommt in den Wortlauten vor, die erklärungsbedürftigen Verhalten von Figuren beschreiben. Es handelt sich gewöhnlich um die Figuren die dem reisenden Helden begegnen und ihn auf die Suche nach der entsprechende Deutung anregen, z. B.:

›F:Folge von Bigotterie deuten‹, ›F:Folge von Geiz deuten‹ etc.

Die Wortlaute mit den Stichworten beheben, erhalten, erfassen, entgegentreten, anpacken, besorgen, herbeiholen, nachheben und deuten können bereits geklonte oder abgeleitete Variante sein.

Ein weiteres, verhältnismäßig wichtiges Teil des Markups besteht aus sog. ›Monowortlaute‹. Sie bezeichnen gewöhnlich Ergebnisse oder Zwischenergebnisse der Handlungen, die durch die geklonten oder abgeleiteten Wortlaute erfasst sind oder den sog. Mehrwert von Aktionen darstellen, die im Laufe der Geschichte keinen Ausklang bekommen Z. B.:

›H:Zur Welt kommen‹, ›H:Heiraten‹, ›H:Bestrafung\_erfolg‹ etc.

Die Methoden wie Klonen und Ableitung der Wortlaute fördern die einheitliche und benutzerfreundliche Anwendung des daraus entwickelten Markups und gewähren eine hohe Qualität der damit annotierten Daten.[[9]](#footnote-11)

### Attribut d

Wie bereits bekannt, kann es im Märchen-Genre insgesamt 24 handlungstragende Figuren geben. In einem Auftritt können gewöhnlich eins bis drei solchen Figuren agieren. Agieren bedeutet, dass die Figur in der Erzählung auftritt bzw. erscheint und nicht das, dass eine solche Figur von einer in der Episode agierende Figur erwähnt, gemeint oder bestellt wird), zu berücksichtigen ist auch der Sachverhalt, dass Figuren mit gleiche Funktionen nur dem Auftritt eines Protagonisten entspricht, z. B. ›zwei älteren Brüder = fHD‹.

Unter der Berücksichtigung der Höhenzahl der Kombination zwischen den Figuren, haben wir darauf verzichten müssen, die Reihenfolge der Figuren nach einer syntaktischen Funktion, wie z.B. das Subjekt / direktes Objekt / indirekte Objekt zu erfassen. Stattdessen haben wir überall die Interaktionen mit unterschiedlich kombinierten und / aber gleichen Figuren auf jeweils eine Kombination reduziert.

(ab|ba) = (ab)

(abc|acb|cab|cba|bac|bca) = (abc)

Dabei sind die handlungstragenden Figuren nach einer festen Reihenfolge (siehe Tab. 1) kombiniert. D.h.: für die zweigliedrige Kombination wird wie folgt verfahren:

* Als erstens wird der ›rHD‹ mit allen anderen kombiniert,
* Zweitens wird der ›rHH‹ ebenfalls mit allen anderen kombiniert, ausgenommen der bereits eingesetzter ›rHD‹,
* Drittens wird das ›rRE‹ mit allen anderen kombiniert, ausgenommen ›rHD‹ und ›rHH‹ usw.

Somit setzt sich die Menge der Variationen von ›d‹ Elemente aus den Untermengen von Einzeln, zu zweit oder zu dritt agierenden Figuren zusammen:[[10]](#footnote-12)

{

*{rHD, rHH, rRE, rHF, rST rVB, rZM, rBZ, rZO, rHP, rPP, rAN, fAN, fPP, fHP, fZO, fBZ ... }*

*{rHD\_rHH, rHD\_rRE, rHD\_rHF, rHD\_ST, rHD\_rVB, rHD\_rZM, rHD\_rZO, rHD\_HP, rHD\_rPP ... }*

*{rHD\_rHH\_rRE, rHD\_rHH\_rHF, rHD\_rHH\_rST, rHD\_rHH\_rVB, rHD\_rHH\_rZM, rHD\_rHH\_rBZ ... }*

}

# Musterbeispiel

Im Folgenden versuchen wir, den Zusammenhang zwischen dem empirischen Objekt und dem darauf basierenden Modell anhand eines Beispiels zu veranschaulichen. Dazu werden wir einen klassischen Märchentext in seine Szenen segmentieren.[[11]](#footnote-13) Jedes Segment wird zusammengefasst und mit dem entsprechenden Markupelement versehen.

|  |  |  |  |
| --- | --- | --- | --- |
| 1. | Auftritt: |  | Der König |
|  | Inhalt: |  | Ein kranker König liegt im Sterben. |
|  | Annotation: | 1 | a551h:KRANKHEIT:rHH |
| 2. | Auftritt: |  | Die drei Söhne und der alte Mann |
|  | Inhalt: |  | Ein weiser alter Mann weist die verzweifelten Königssöhne auf das schwer zu beschaffende Lebenswasser hin. |
|  | Annotation: | 2 | a551: HF:HEILMITTEL\_besorgen:rHD\_rVB\_fHD |
| 3. | Auftritt: |  | Der Vater und der älteste Sohn |
|  | Inhalt: |  | Der älteste Sohn bittet den Vater um Erlaubnis und macht sich auf den Weg, das Lebenswasser zu finden. |
|  | Annotation: | 3 | a551:F: Werben\_um\_den\_Auftrag:rHH\_fHD |
| 4. | Auftritt: |  | Der älteste Sohn und der Zwerg |
|  | Inhalt: |  | Der ältere Sohn zeigt Arroganz gegenüber dem Zwerg. Dieser wird zornig auf den ältesten Sohn und sorgt dafür, dass er gelähmt wird. |
|  | Annotation: | 4  5  6 | a551:F:Gutes\_Benehmen:rST\_fHD  a551:h:Gutes\_Benehmen:rST\_fHD  a551:h: FREIHEITSBERAUBUNG:rST\_fHD |
| 5. | Auftritt: |  | Der Vater und der mittlere Sohn |
|  | Inhalt: |  | Der mittlere Sohn bittet den Vater um Erlaubnis und begibt sich auf die Suche nach Lebenswasser. |
|  | Annotation: | 7 | a551:F: Werben\_um\_den\_Auftrag:rHH\_fHD |
| 6. | Auftritt: |  | Der mittlere Sohn und der Zwerg |
|  | Inhalt: |  | Der mittlere Sohn zeigt ebenfalls Arroganz gegenüber dem Zwerg. Dieser wird wütend auf den mittleren Sohn und sorgt dafür, dass er gelähmt wird. |
|  | Annotation: | 8  9  10 | a551:F: Gutes\_Benehmen:rST\_fHD  a551:h: Gutes\_Benehmen:rST\_fHD  a551:h: FREIHEITSBERAUBUNG:rST\_fHD |
| 7. | Auftritt: |  | Der Vater und der jüngste Sohn |
|  | Inhalt: |  | Der jüngste Sohn bittet den Vater um Erlaubnis und macht sich auf den Weg, Lebenswasser zu suchen. |
|  | Annotation: | 11 | a551:F:Werben\_um\_den\_Auftrag:rHH\_fHD |
| 8. | Auftritt: |  | Der jüngste Sohn und der Zwerg |
|  | Inhalt: |  | Der jüngste Sohn verhält sich höflich gegenüber dem Zwerg. Dieser gibt dem jüngsten Sohn notwendige Anweisungen. |
|  | Annotation: | 12  13  14 | a551:F:Gutes\_Benehmen:rHD\_rST  a551:H:Gutes\_Benehmen: rHD\_rST  a551:HF:PFÖRTNER\_besänftigen:rHD\_rST |
| 9. | Auftritt: |  | Der jüngste Sohn und die Wache |
|  | Inhalt: |  | Der jüngste Sohn besänftigt die Wache, die ihn passieren lässt. |
|  | Annotation: | 15  16 | a551:H: PFÖRTNER\_besänftigen:rHD\_fHF  a551:H: PFÖRTNER\_entkommen:rHD\_fHF |
| 10. | Auftritt: |  | Der jüngste Sohn und die verwunschenen Wesen |
|  | Inhalt: |  | Der jüngste Sohn ist der Zaubergegenstände handhabt geworden. Die Personen von denen er diese Gegenstände erhält sind nicht zu kategorisieren. |
|  | Annotation: | 17 | a551:H:ZAUBERMITTEL\_erhalten:rHD\_rZM |
| 11. | Auftritt: |  | Der jüngste Sohn und die Besitzerin des Lebenswassers |
|  | Inhalt: |  | Es entwickelt sich eine Liebesbeziehung zwischen dem jüngsten Sohn und der Besitzerin des Lebenswassers. Der muss aber die Schöne bald verlassen. |
|  | Annotation: | 18  19  20 | a551:H:BESITZER\_DES\_OBJEKTES\_erfassen:rHD\_fBZ  a551:H:LIEBE\_wird\_ausgelöst:rHD\_fBZ  a551:h:KONTAKR\_abbrechen: rHD\_fBZ |
| 12. | Auftritt: |  | Der jüngste Sohn und das Lebenswasser |
|  | Inhalt: |  | Der jüngste Sohn erlangt das Lebenswasser. |
|  | Annotation: | 21 | a551:H:HEILMITTEL\_erfassen:rHD\_rZO |
| 13. | Auftritt: |  | Der jüngste Sohn und der Zwerg |
|  | Inhalt: |  | Der jüngste Sohn bittet den Zwerg um die Freilassung seiner Brüder, ignoriert dabei aber dessen Warnung. |
|  | Annotation: | 22  23  24 | a551:F:Auf\_Hilfsquelle\_zurückgreifen:rHD\_rST  a551:HF:ABRATEN\_von\_derMission:rHD\_rST  a551:h:ABRATEN\_von\_der\_Mission\_nachgehen:rHD\_rST |
| 14. | Auftritt: |  | Der jüngste Sohn und seine Brüder |
|  | Inhalt: |  | Der jüngste Sohn trifft seine befreiten Brüder und teilt ihnen von seinem Erfolg mit. |
|  | Annotation: | 25 | a551:Hh:FREIHEITSBERAUBUNG\_beheben:rHD\_fHD |
| 15. | Auftritt: |  | Die Könige aus fernen Ländern und der jüngste Sohn |
|  | Inhalt: |  | Auf dem Rückweg nach Hause unterstützt der jüngste Bruder die Herrscher von drei unterschiedlichen Ländern mithilfe seiner magischen Gegenstände. |
|  | Annotation: | 26  27 | a551:F:Sorge\_um\_Hilfsbedürftige:rHD\_rHF\_rZM  a551:H:Sorge\_um\_Hilfsbedürftige:rHD\_rHF\_rZM |
| 16. | Auftritt: |  | Die Könige aus fernen Ländern und der jüngste Sohn |
|  | Inhalt: |  | Auf dem Rückweg nach Hause unterstützt der jüngste Bruder die Herrscher von drei unterschiedlichen Ländern mithilfe seiner magischen Gegenstände. |
|  | Annotation: | 28  29 | a551:F:Sorge\_um\_Hilfsbedürftige:rHD\_rHF\_rZM  a551:H: Sorge\_um\_Hilfsbedürftige:rHD\_rHF\_rZM |
| 17. | Auftritt: |  | Die älteren Söhne beim Schlaf ihres jüngsten Bruders |
|  | Inhalt: |  | Die älteren Brüder werden neidisch und tauschen heimlich das Lebenswasser gegen gewöhnliches Wasser aus. |
|  | Annotation: | 30  31 | A551:f:MISSETAT\_planen:rZM\_fHD  a551:h:MISSETAT\_wie\_Aneignung:rZM\_fHD |
| 18. | Auftritt: |  | Der jüngste Sohn und ihr Vater |
|  | Inhalt: |  | Das beschaffte Heilmittel macht den König noch kränker. |
|  | Annotation: | 32  33 | a551: h:HEILMITTEL\_besorgen: :rHD\_rHH\_fZO  a551:h:KRANKHEIT\_beheben:rHD\_rHH\_fZO |
| 19. | Auftritt: |  | Die älteren Söhne der Vater und das Heilmittel |
|  | Inhalt: |  | Die älteren Söhne heilen den Vater und verleumden den jüngsten Bruder. |
|  | Annotation: | 34  35  36 | a551:h:HEILMITTEL\_besorgen: : rHH\_rZO\_fHD  a551:Hh:KRANKHEIT\_beheben: rHH\_rZO\_fHD  a551:h:VERLEUMDUNG\_bei\_Untreue: rHH\_rZO\_fHD |
| 20. | Auftritt: |  | Die älteren und der jüngste Bruder |
|  | Inhalt: |  | Die älteren Brüder bringen den jüngsten Bruder zum Schweigen. |
|  | Annotation: | 37 | a551:h:SCHWEIGEPFLICHT\_verhängen:rHD\_fHD |
| 21. | Auftritt: |  | Der König und der Scharfrichter |
|  | Inhalt: |  | Der König befiehlt den Tod seines jüngsten Sohns. |
|  | Annotation: | 38 | a551:f:TODESAUFTRAG:rVB\_rHH |
| 22. | Auftritt: |  | Der ältere Sohn und der Scharfrichter |
|  | Inhalt: |  | Der Scharfrichter weigert sich, den Prinzen zu töten, und dieser flieht. |
|  | Annotation: | 39  40 | A551:H:Mitleidsgefühl\_auslösen:rHD\_rVB  a551:H: TODESAUFTRAG\_entkommen:rHD\_rVB |
| 23. | Auftritt: |  | Der Vater, die drei fremden Könige |
|  | Inhalt: |  | Die drei Könige, denen der Prinz einst half, besuchen das Königreich. Der König erfährt die Wahrheit und vergibt seinem jüngsten Sohn. |
|  | Annotation: | 41  42 | a551:H:Auf\_Hilfquelle\_zurückgreifen:rHH\_rHF  a551:H: VERLEUMDUND\_entgegensetzen: rHH\_ rHF |
| 24. | Auftritt: |  | Der König und der Scharfrichter |
|  | Inhalt: |  | Der König erfährt vom Scharfrichter, dass sein jüngster Sohn freigelassen wurde, und erkennt die Verleumdung. |
|  | Annotation: | 44 | a551: H:VERLEUMDUNG\_beheben:rHH\_rVB |
| 25. | Auftritt: |  | Die Besitzerin des Zaubermittels |
|  | Inhalt: |  | Gleichzeitig bereitet sich die Besitzerin des Lebenswassers auf die Suche nach ihrem Geliebten vor. |
|  | Annotation: | 45 | a551:F:KONTAKT\_wiederherstellen:fBZ |
| 26. | Auftritt: |  | Der älteste Sohn und der Diener der Besitzerin des Lebenswassers |
|  | Inhalt: |  | Der älteste Sohn scheitert ebenfalls beim Identitätstest. |
|  | Annotation: | 46 | a551:h:IDENTITÄT\_beweisen:fHD |
| 27. | Auftritt: |  | Der mittlere Sohn und der Diener der Besitzerin des Lebenswassers |
|  | Inhalt: |  | Der mittlere Sohn scheitert ebenfalls beim Identitätstest. |
|  | Annotation: | 47 | a551:h:IDENTITÄT\_beweisen:fHD |
| 28. | Auftritt: |  | Der jüngste Sohn und die Besitzerin des Lebenswassers |
|  | Inhalt: |  | Der jüngste Sohn besteht den Test und wird erkannt. |
|  | Annotation: | 48  49 | a551:H:IDENTITÄT\_beweisen:rHD\_fBZ  a551:H:KONTAKT\_wiederherstellen:rHD\_fBZ |
| 29. | Auftritt: |  | Vater und der älteste Sohn |
|  | Inhalt: |  | Der älteste Sohn gesteht dem König die Wahrheit. Die Verbrecher fliehen und kehren nie zurück. |
|  | Annotation: | 50  51 | a551:H: SCHWEIGENPFLICHT\_beheben:rHD\_rHH\_fHD  a551:H:Bestrafung\_erfolgt:rHD\_rHH\_fHD |

# Motiv, Episode, Märchen

Die Organisation der inhaltlichen Elemente im Märchen sind auf zwei Strukturebene empirisch zu beobachten: das sind die Motivebene und der Ebene der Vollgeschichte. Das erste lässt sich durch den Anfang und Ende des Auftritts in der Geschichte erkennen, das andere durch den Anfang und Ende des gesamten Textes.

Die Vollgeschichte als oberste Instanz der Analyse entspricht einer Makrostruktur, die aus den in den mesoskopischen Strukturebene organisierten Mikroelementen besteht. Bei der Annahme dieser Definition für die Vollgeschichte, ist das Motiv hingegen als kleinstes und somit mikrostrukturelles Element des Systems zu definieren. Was zwischen der Vollgeschichte und dem Motiv vorkommt, gehört zum mesoskopischen Strukturelement. Als solches sind die Episoden zu zählen.

Die mesoskopische Natur der Episode beinhaltet automatisch die Repräsentationsfähigkeit der Charakter von Mikro– und Makroelemente. Daher ergibt sich die Möglichkeit, dass eine Episode sowohl ein Baustein als auch ein Gebilde sein kann, das selbst aus solchen Bausteinen besteht.

Anhand unseres Musterbeispiels versuchen wir den Zusammenhang zwischen den Strukturelementen ›Motiv‹, ›Episode‹ und ›Märchen‹ zu präsentieren. Dafür werden wir die Annotationen als Motivliste darstellen (siehe die Werte ›Annotation‹ in der Tab. 4) und die Zusammenhänge der Motive in Episoden erfassen.

In einer Episode wird die Geschichte einer großen oder kleinen Krise erzählt. Die Episode folgt einer Struktur, die aus der Offenkundigkeit, Handlung und Lösung der Vorhandenen oder Auslösung einer neuen Krise besteht.

Wie bereits bekannt, wird die Behebung der Krise durch das Motiv mit den Attributwerten ›b‹ = ›H‹ / ›Hh‹ dargestellt. Die Auslösung oder Offenkundigkeit der Krise wird hingegen durch ein Motiv erfasst, das dem letzten vorabgeht, und einen vergleichbaren Wortlaut (Attributwert ›c‹) hat[[12]](#footnote-14). Diese Verbindung zwischen den Motiven qualifizieren wir als Kausalität ersten Grades.

Andere Motive, die sich auf dieselbe Krise beziehen, sollen sich in unmittelbarer Nähe zu den oben genannten Motiven befinden und gleiche ›d‹ Attributwert haben. Diese Verbindung nennen wir die Kausalität des zweiten Grades.

Zur niedrigsten, dritten Stufe der Kausalität gehören die mit den Motiven der letzten Kategorie angehängte Elemente mit vergleichbaren ›c‹ Attributwerten (geklont oder abgeleitet) sowie die Elemente mit identischen ›d‹ Attributwerte, Ausnahme gilt für den Helfer (›HF‹) bzw. Stifter (›ST‹) dessen Anwesenheit allein oder in der Kombination mit beliebiger anderer Figur (›d‹ Attributwerten) auf solche Kausalität hinweist.

Die Feststellung der Kausalität zwischen den Motiven einer Episode bzw. Episodencluster läuft von unten nach oben in die gesamte Motivliste. Dem letzten Motiv mit dem Attributwert ›b‹ = ›H‹ / ›Hh‹ soll in oberen Bereich der Liste ein Motiv entsprechen, das einen vergleichbare ›c‹ Attributwert hat.

Als nächster Schritt sollen die Motive erfasst werden, die in der unmittelbaren Nähe von solchen Motiven stehen und gleiche ›d‹ Werte haben.

Die Tatsache, dass durch diese Recherche komplexe Episodencluster herausgetragen werden oder einige Motive gleichzeitig in mehreren Episoden auftauchen, ist vollkommen normal und weist auf den Sachverhalt hin, demzufolge der konsekutive Abschluss einer Krise und somit auch einer Episode den Beginn einer neuen Krise bzw. Episode bedeuten kann.

Folgendes Beispiel:

Der Held steht vor der Wahl, gegenüber dem Zwerg entweder genau so frech zu verhalten wie seine älteren Brüder ›h:Gutes\_benehmen:rST\_fHD‹ oder im Gegensatz zu denen, sich höflich benehmen. Der Held verhält sich höflich, also wählt positive Lösung der Krise aus ›H:Gutes\_benehmen:rHD\_rST‹. Als Mehrwert der Lösung erhält er Handlungsanweisung darüber, wie er die Pförtner besänftigen kann, damit er einen freien Zugang zum Zielobjekt beschafft. Das Motiv lautet wie folgt: ›HF:PFÖRTNER\_besänftigen:rHD\_rST‹.

Der mit der Handlungsanweisungen des Stifters (hier des Zwerges) ausgerüstete Held gewinnt die Sympathie der leidende Wache und verschafft einen freien Zugang zum Zielobjekt (hier zum Heilmittel):

›H: PFÖRTNER\_besänftigen :rHD\_fHF‹, ›H: PFÖRTNER\_besänftigen:rHD\_fHF‹.

Hier ein Beispiel der nach dieser Regel festgestellte Episode anhand der Motivliste des oben annotierten Musterbeispiels:

1. Beispiel: Gewinnung der Schönen (Besitzer des Zielobjektes hier Besitzerin des Lebenswassers)

›Kausalität der ersten Grad‹

›Kausalität der zweiten Grad‹

› Kausalität der dritten Grad‹

H:Die\_SCHÖNE\_erfassen:rHD\_fBZ (18)

H:LIEBE\_wird\_ausgelöst:rHD\_fBZ (19)

h:KONTAKT\_abbrechen:rHD\_fBZ (20)

F:KONTAKT\_wiederherstellen: fBZ (45)

h:IDENTITÄT\_beweisen:fHD\_fVB (46)

h:IDENTITÄT\_beweisen:fHD\_fVB (47)

H:IDENTITÄT\_beweisen:rHD\_fBZ (48)

H:KONTAKT\_wiederherstellen:rHD\_fBZ (49)

2. Beispiel: Auslösung von Krankheit und deren Beseitigung

Beispiel: Auslösung von Krankheit und deren Beseitigung

›Kausalität der ersten Grad‹

›Kausalität der zweiten Grad‹

› Kausalität der dritten Grad‹

h:KRANKHEIT:rHH (1)

F:HEILMITTEL\_besorgen:rHD\_rVB\_fHD (2)

F:Sorge\_um\_Hilfsbedürftige:rHD\_rHF\_rZM

H: Sorge\_um\_Hilfsbedürftige:rHD\_rHF\_rZM

F: Sorge\_um\_Hilfsbedürftige:rHD\_rHF\_rZM

H: Sorge\_um\_Hilfsbedürftige:rHD\_rHF\_rZM

h:HEILMITTEL\_besorgen:rHD\_rHH\_fZO (32)

h:KRANKHEIT\_beheben:rHD\_rHH\_fZO (33)

Hh:HEILMITTEL\_besorgen:rHH\_rZO\_fHD (34)

Hh:KRANKHEIT\_beheben:rHH\_rZO\_fHD (35)

h:VERLEUMDUNG\_von\_Untreue:rHH\_rZO\_fHD (36)

H:Auf\_Hilfsquelle\_zurückgreifen:rHH\_rHF (41)

H:VERLEUMDUNG\_entgegentreten:rHH\_rHF (42)

H:VERLEUMDUNG\_beheben:rHH\_rHF (44)

Der Aufgabe, wie diese Regel Zwecks der vollautomatischen Analyse des Märchens eingesetzt werden kann ist noch zu erledigen.

## **Maschinelle Analyse des Märchens**

### **Einführung**

Angesichts ihrer einfachen Formen und Strukturen sind Genres der mündlichen Überlieferung beliebte Objekte zur Demonstration der Effektivität analytischer und generativer Modelle der künstlichen Intelligenz. Viele dieser Versuche basieren auf dem wissenschaftlichen Erbe der Finnischen Schule, insbesondere auf den ATU– und MIT–Klassifikationssystemen.[[13]](#footnote-15) Sie beschränken sich jedoch darauf, diese Instrumentarien zu automatisieren, ohne ihre Grundmethoden nach den Anforderungen des neuen Ansatzes, den Methoden der digitalen Geisteswissenschaften, grundlegend zu reformieren.[[14]](#footnote-16) Es ist weiterhin populär, die durch die strukturalistische Analyse des Märchens erfassten Textbausteine, die sogenannten Propp'schen Funktionen, zu formalisieren und ihre maschinelle Erkennbarkeit zu prüfen.[[15]](#footnote-17)

Im Gegensatz zu diesen Versuchen, die als Retrodigitalisierung der traditionellen Methoden rekapituliert werden können, konzentriert sich die vorliegende Arbeit auf die Entwicklung einer qualitativ neuen, auf digitalen Ansätzen basierenden Methode zur Lösung traditioneller Fragestellungen der vergleichenden Märchenforschung.

Die wichtigste Aufgabe dieses neuen Ansatzes besteht in der formalisierten Erschließung des Forschungsobjekts und nicht in der Anpassung vorhandener Erkenntnisse und Systeme an digitale Anwendungen oder umgekehrt.

Um den artikulierten Unterschied zwischen der auf dieser Strategie beruhenden Forschung und den Ergebnissen der bereits veröffentlichten Abhandlungen im Bereich der elektronischen Märchenforschung zu verdeutlichen, versuchen wir hier beispielhaft, eine der jüngsten Publikationen zu diesem Thema zu besprechen. Es handelt sich um den Aufsatz von J. Eklund und anderen unter dem Titel: *"Teaching Tale Types to a Computer: A First Experiment with the Annotated Folktales Collection"* [[16]](#footnote-18).

Wie aus dem genannten Artikel hervorgeht, setzen sich die Autoren zum Ziel, den Computer dazu zu bringen, die Zugehörigkeit der Texte zu den bekannten Märchentypen vorherzusagen. Als Basis für die Klassifikation dient der internationale Märchenkatalog (ATU). Die maschinelle Erkennung des Typs beruht auf den inhaltlichen Eigenschaften des Textes, die wiederum durch die im Motivkatalog von Thompson (MIT) erfassten inhaltlichen Einheiten identifiziert werden. Laut dem Aufsatz haben die Autoren den Computer in die Lage versetzt, die MIT-Elemente im Text zu erkennen und je nach deren Kombination im Text vorherzusagen, welchem Typ der Text zuzuordnen ist.

Zunächst etwas Näheres zu den Katalogen ATU und MIT. Der erste Versuch einer Katalogisierung von inhaltlichen Übereinstimmungen in den Märchenrepertoires geht auf die Vertreter der mythologischen Schule zurück. Es handelt sich um die handschriftlichen Sagenkonkordanzen der Brüder Jacob und Wilhelm Grimm. Leider haben die Arbeiten der Brüder nicht den Stand der Veröffentlichungsreife erreicht.[[17]](#footnote-19)

Die von den Brüdern Grimm erkannte Notwendigkeit der Katalogisierung von Übereinstimmungen verschiedener Märchenrepertoires wurde in relativ kleinem Maßstab von ihrem Anhänger J. G. von Hahn umgesetzt. Der von ihm in seiner griechischen und albanischen Märchensammlung angefügte Formelkatalog gilt als erstes Verzeichnis der Märchentypen[[18]](#footnote-20).

Die Finnische Schule legte großen Wert auf die Katalogisierung des Märchens. Antti Aarne erklärte die Notwendigkeit, einen ›systematischen Typenkatalog auszuarbeiten‹, aufgrund des Mangels an einem System, das in der Vielfalt der Volksmärchen die einzelnen Märchentypen feststellen und zu einem geordneten Ganzen vereinen könnte[[19]](#footnote-21). Damit begann die Finnische Schule bzw. die vergleichende Märchenforschung, über stabile Inhaltselemente zu sprechen und dabei mit entsprechenden wissenschaftlichen Begriffen zu operieren.

Die Klassifikation nach Typen gewährleistet, dass ähnliche Geschichten durch gemeinsame Identifikationsmerkmale gekennzeichnet und somit miteinander vergleichbar sind. Von Anfang an wurde der Katalog ausschließlich für angewandte Zwecke eingeführt. Die Implementierung und die kontinuierliche Pflege des Klassifikationssystems führten jedoch dazu, dass der Katalog allmählich von einem rein angewandten zu einem wissenschaftlichen System wurde. der Vorgang und die Konsequenzen dieser Entwicklung lassen sich wie folgt beschreiben:

1. Durch empirische Erkenntnisse über die stabil wiederholbare Inhaltselemente in Märchen aus unterschiedlichen Kulturen entstand der Bedarf, deren Vergleichbarkeit zu katalogisieren.

2. Die Katalogisierung erfüllt eine angewandte Funktion und zielt darauf ab, die Aufgabe eines Nachschlagewerks zu erfüllen.

3. Nationale Märchenkataloge identifizieren und klassifizieren Texte nach diesem System in ihren nationalen Repertoires.

4. Nach dieser Phase bietet der Katalog zusätzliche und erkenntnisrelevante Informationen für die Forschung, um folgende Fragen zu beantworten:

(a) Wie sieht die Rangliste der Typen im nationalen Märchenrepertoire aus?

(b) Aus welchen Variationen bestehen einzelne Typen in diesen Repertoires?

5. Später wird versucht, die Typenbeschreibungen aus verschiedenen lokalen Katalogen zusammenzuführen, um verallgemeinerte Zusammenfassungen einzelner Typen darzustellen.

Der Versuch, sowohl spezifische als auch allgemeine Informationen in einem Internationalen Katalog des Märchens zusammenzuführen, ist an der Komplexität der dafür kodierten Informationen gescheitert. Der Katalog identifizierte zum Beispiel neue Fälle von Kombinationen oder Kontaminationen, konnte jedoch nicht gleichzeitig angeben, aus welchen Repertoires diese Fälle stammen oder ob sie dort typische oder eher zufällige Einzelfälle darstellen.

Dieser Versuch geht auf den amerikanischen Erzählforscher Thompson zurück. Er entwickelte außerdem den Index von internationalen Motiven (MIT) und integrierte diesen in den Typenkatalog, um ihn seiner Meinung nach in ein noch präziseres Informationssystem umzuwandeln. Ab jetzt sind neben den traditionellen Zusammenfassungen des Märchens auch die Indizien zu finden, die auf die festen Inhaltselemente, die sogenannten Motive, hinweisen. Zum Beispiel sind bei der Beschreibung des Typs ›ATU 300 The Dragon-Slayer‹ folgende Indizes zu finden:

›B421 = Helpful dog‹.

›B312.2. Helpful animals obtained by exchange‹.

›B11.2.3. 1. Seven-headed dragon‹.

›B1 1.10. Sacrifice of human being to dragon‹.

›S262. Periodic sacrifices to a monster‹.

›T68.1. Princess offered as prize to rescuer‹.

›D1975. Dragon-fighter's magic sleep‹.

›D1978.2. Waking from magic sleep by letting tear fall on sleeper‹.

Diese stammen aus dem internationalen Motivindex (MIT), wo sie nach dem Prinzip der Taxonomie in verschiedene Themen und Unterthemen eingeteilt sind. Zum Beispiel:

»B = Animal Motifs.

B11= Dragon.

B11.2 = Form of dragon.

B11.2.3 = Many-headed dragon.

Die danach kommende Zahl weist auf die Menge der Drachenköpfe hin:

B11.2.3. 1 = Seven-headed dragon.

B11.2.3.2 = Three-headed dragon.

B11.2.3.3 = Six-headed dragon.

usw.«[[20]](#footnote-22)

Nach der Integration der MIT-Systemelemente in das ATU-System wurden die traditionellen Typenzusammenfassungen um zusätzliche MIT-Indizes ergänzt. Ein Beispiel dafür ist die folgende Zusammenfassung des oben genannten Typs:

»ATU 300 The Dragon-Slayer. A youth (e.g. by exchange three wonderful dogs [B421, B312.2]. He comes to a town where people are mourning and learns that once a year a (seven-headed) dragon [B11.2.3.1] demands a virgin as a sacrifice [B11.10, S262]. In the current year, the king’s daughter has been chosen to be sacrificed, and the king offers her as a prize to her rescuer [T68.1]. The youth goes to the appointed place. While waiting to fight with the dragon, he falls into a magic sleep [D1975], during which the princess twists a ring (ribbons) into his hair; only one of her falling tears can awaken him [D1978.2]. usw.«[[21]](#footnote-23)

Nach Ansicht der Autoren des betrachteten Aufsatzes gilt folgendes: Wenn diese MIT-Elemente im Text in einer mehr oder weniger ähnlichen Reihenfolge vorkommen, kann man daraus schließen, dass es sich um den Typ ATU 300 handelt. Durch das Training des Modells mit entsprechend gelabelten Trainingsdaten kann es dann die Fähigkeit erlangen, die Zugehörigkeit des Textes zum passenden Typ vorherzusagen.

Die Effizienz eines solchen Unternehmens ist nicht zu unterschätzen. Dennoch stellt sich die Frage, ob wir dadurch tatsächlich mehr über die Unterschiede und Ähnlichkeiten zwischen den einzelnen Texten oder Repertoires erfahren können. Dies hängt vollständig davon ab, wie adäquat wir das Objekt formalisieren können und nicht davon, wie geschickt wir den Computer bei der Anwendung der kompromissartigen Arbeitsmethode einsetzen.

Wenn wir einen Text erschließen müssen und dabei über keinen Computer, sondern nur ein Zettelsystem verfügen, müssen wir so viele Zettel erstellen, wie es Lemmata im Text gibt, multipliziert mit der Anzahl der Sätze, in denen diese Lemmata ein- oder mehrmals vorkommen. Wenn wir jedoch nach der Erledigung dieser Aufgabe einen Computer zur Verfügung haben, müssen wir nicht die handschriftlich ausgefüllten Zettel als einzelne Datensätze in den Computer eintragen, sondern können den fließenden Text eingeben und diesen später je nach den Fragestellungen auswerten.

Genau dasselbe gilt für die Wiedergeburt der Arbeitsmethode der vergleichenden Märchenforschung. Die Idee, Märchen durch eine gemeinsame Instanz (tertium comparationis) zu vergleichen, bleibt dabei nach wie vor relevant. Welches Modell wir für die Konstruktion des Typs als Vergleichsobjekt verwenden, muss allerdings aus der Perspektive computergestützter Lösungsstrategien bestimmt werden.

Eine automatisierte Erkennung des Textes ohne kritische Auseinandersetzung mit dem Erbe der vergleichenden Märchenforschung aus der Zeit vor den digitalen Geisteswissenschaften kann nichts anderes sein als eine Retrodigitalisierung einer Methode, die nicht nach den Anforderungen der formalisierten Erschließung des Forschungsobjektes erarbeitet wurde. Wie bereits angedeutet, distanzieren wir uns von dieser Herangehensweise und versuchen, die Schwächen der bisherigen Methode der vergleichenden Märchenforschung durch eine neue Lösung des traditionell erkannten Forschungsproblems mithilfe computergestützter Methoden zu bewältigen.

Der Grund für die Entstehung des vorliegenden Vorhabens liegt im Bedarf an einem auf philologischer Analyse basierenden Modell, das für die maschinelle Analyse von Märchentexten verwendet werden kann. Die maschinelle Analyse des Forschungsobjektes anhand der formalisierten Beschreibung der Struktur des Märchens beinhaltet sowohl die Automatisierung der Segmentierung des Märchens in Strukturelemente unterschiedlicher Größe als auch das Erzeugen eines zusammenfassenden Bildes aufgrund des Vergleichs der je nach Fragestellung ausgewählten Forschungsobjekte. Beispiele hierfür sind:

(a) Die vergleichende Analyse des Typs X anhand der Repertoires A, B, C usw.

(b) Die Kombination von Motiven innerhalb der Episode X.

(c) Der der Status der handlungstragenden Figuren im Motiv Y.

Und vieles mehr.

Die Lösung der maschinellen Analyse des Märchens als nächster Schritt nach der Formalisierung des Forschungsobjektes ist mit der Entwicklung der entsprechenden digitalen Infrastruktur verbunden. Im Folgenden werden wir einige der wichtigsten Elemente dieser Infrastruktur betrachten.

### **Textkorpus**

Das Textkorpus ist eine XML-Datei, die auf dem TEI-Standard basiert und das Wurzelelement ›teiCorpus‹ hat.[[22]](#footnote-24) Die Texte innerhalb des Korpus können in verschiedenen Sprachen vorliegen, müssen jedoch gemäß den Open-Data-Anforderungen standardisiert und in Unicode kodiert sein. Der gezielte Zugriff auf die Texte aus verschiedenen Sprachen oder Gruppen kann anhand der im ›TEI‹-Element erfassten Metadaten erfolgen. Hierbei ist die Anwendung des ›xml:id‹ Attributs beim Element ›text‹ empfehlenswert. Dieses Attribut besteht aus vier Teilen, die Informationen über die Lizenzbedingungen[[23]](#footnote-25), über die Zugehörigkeit des Textes zu einem bestimmten Repertoire, die Sprache des Textes und die laufende Nummer des Textes unter anderen Materialien mit gleicher Zugehörigkeit zum Repertoire enthalten.

Folgende Beispiele

›cc\_deu\_deu\_1‹ = 1. Text aus dem deutschen Repertoire in deutsche Sprache

›cc\_deu\_eng\_1‹ = 1. Text aus dem deutschen Repertoire in englische Sprache

›cc\_deu\_eng\_2‹ = 2. Text aus dem deutschen Repertoire in englische Sprache usw.

Für die Benennung der Namen der Repertoires und Sprachen werden Abkürzungen gemäß dem ISO-639-3-Standard verwendet. Die Abkürzung cc am Anfang des Attributwertes weist auf die Freinutzungsmöglichkeit der Datei hin.

Im TEI-Korpus sind die Märchen einzeln in separaten TEI-Elementen erfasst. Für jedes Märchen liegt eine codierte oder zur Codierung bereitgestellte Textfassung vor. Die Basistexte befinden sich in verschiedenen Repositorien und sind untereinander sowie mit dem codierten Text entsprechend referenziert. Die Texte, die nicht den OpenData-Forderungen entsprechen, werden im Textkorpus in abgeleiteter Form dargestellt und mit bibliographischen Angaben zur primären Textquelle versehen.

Die Hauptaufgabe des Korpus besteht darin, eine standardisierte Segmentierung der Daten sowie einen gezielten Zugriff auf diese Daten zu gewährleisten. Das entwickelte Datenmodell ermöglicht es, Auftritte im Text auseinander zu differenzieren und in ihr eingebettete Motive zu erfassen. Dafür sind das Element ›seg‹ und in ihm eingebettete Tupel von Attributen ›a‹, ›b‹, ›c‹ und ›d‹ zuständig. Die optimale Anzahl von Motiven innerhalb eines Auftritts beträgt fünf. Dies bedeutet, dass jedes Motiv besteht aus zwei eigenen Attributen ›b‹ und ›c‹ sowie weitere zwei gemeinsame Attribute ›a‹ und ›d‹ besteht. Wie bereits bekannt, liegt der Segmentation des Textes der neue Auftritt einer oder mehrere handlungstragenden Personen / Personen zugrunde, dementsprechend alles, was bei diesem Auftritt passiert, d. h. theoretisch alle fünf ›b‹ und ›c‹ Werte, kann nur ein einziges ›d‹ Element (d.h. handlungstragende Person bzw. Personen) haben.

Ein Beispiel für die Annotation des 10. Abschnitts in unserem Text wäre wie folgt (N gilt für die Null):

<seg a:ana = “a551“ b1:ana = “F“ c1:ana = “ Gutes\_Benehmen“ b2:ana = “H“ c2:ana = “ Gutes\_Benehmen “ b3:ana=“HF“ c3:ana=“ Pförtner\_besänftigen “ b4:ana=“N“ c4:ana=“ N“ b5:ana=“N“ c5:ana=“ N“ d:ana=“ rHD\_rHF“ > Als auch der zweite Sohn ausblieb, so erbot sich der jüngste, auszuziehen und das Wasser zu holen, und der König mußte ihn endlich ziehen lassen. Als er dem Zwerg begegnete und dieser fragte, wohin er so eilig wolle, so hielt er an, gab ihm Rede und Antwort und sagte »ich suche das Wasser des Lebens, denn mein Vater ist sterbenskrank.« »Weißt du auch, wo das zu finden ist?« »Nein,« sagte der Prinz. »Weil du dich betragen hast, wie sichs geziemt, nicht überütig wie deine falschen Brüder, so will ich dir Auskunft geben und dir sagen, wie du zu dem Wasser des Lebens gelangst…. </seg>

### **Künstlicher Assistent für die semiautomatische Annotation der Texte**

Ein zentraler Bestandteil der maschinellen Analyse von Märchen ist ein künstlicher Assistent mit Vorhersagefunktion.[[24]](#footnote-26) Dieser Assistent ist dafür zuständig, in einer umfangreichen Menge von Daten die Textabschnitte zu erkennen, die über die gesuchten inhaltlichen Eigenschaften (Themen / Topic) verfügen.

Die Entwicklung eines solchen Assistenten erfordert die Erfüllung der folgenden drei Aufgaben:

1. Auswahl einer geeigneten Vorgehensweise für die maschinelle Textanalyse.

2. Anpassung dieser Vorgehensweise an die Forschungsfrage.

3. Umformulierung des gewonnenen Wissens über das Forschungsobjekt in die Sprache der ausgewählten Vorgehensweise.

Besonders wichtig bei der letzten Aufgabe ist die Anpassung zwischen den Elemente des konzeptuellen Modells und empirisch beobachtbaren Objekte. Ein Auftritt der handlungstragenden Person innerhalb der Geschichte wird dabei als empirisch beobachtbare Grenze für den Anfang bzw. das Ende eines Motivs betrachtet. Für die Maschine bedeutet dies eine Zeichenkette, bei der das kleinste erkennbare Element ein Wort und das größte ein Absatz ist. Auf inhaltlicher Ebene entspricht dies dem Textabschnitt mit veränderten Figurennamen, ihren Attributen, Ortschaften und Handlungen. Wir können zwar der Maschine beibringen, diesen Themenwechsel zu erkennen, jedoch müssen wir zuvor deutlich machen, welcher Zusammenhang zwischen den Roh-, Trainings-, Test- und Zieldaten besteht.

Der rohe Text, wie er uns gewöhnlich vorliegt, ist durch die folgenden explizit erkennbaren Zeichen wahrnehmbar: Wort, Satz, Absatz und Text. Die Grenzen der somit organisierten Textteile entsprechen jedoch nicht unbedingt dem von uns gesuchten Wechsel zwischen den Auftritten der handelnden Personen im Text.

Die Aufbereitung des rohen Textes zu Trainingsdaten erfordert die Umwandlung der sogenannten ›Istzeichen‹ in ›Sollzeichen‹. Dabei ist es oft notwendig, die Anzahl der Absätze im Text zu korrigieren. Die Absätze können entweder zusammengelegt oder in mehrere Teile aufgeteilt werden. In manchen Fällen kann auch ein einziger Satz einem solchen Absatz entsprechen.

Messdaten sind den Rohdaten gleich, sie enthalten bibliographische Metadaten und stellen wohlgeformte XML–Dateien dar. Zieldaten sind annotierten Messdaten. Diese können teilweise oder vollständig annotiert werden. Die bereits annotierten Teile der Messdaten werden bei der automatischen Analyse der Texte nicht miteinbezogen. Sie können allerdings als Trainingsdaten verwendet werden.

Für die automatische Erkennung der Motive und Episoden im Text verwenden wir die Topic Modeling Methode basierend auf den Algorithmus der logistischen Regression (LogisticRegression). Ein Modell, das von diesem Algorithmus gesteuert wird, versucht, Merkmale (Themen) zu finden, die für die Erkennung der gesuchten Episoden (positiv etikettierte Datensätze) sowie ihrer Gegenteile (negativ etikettierte Datensätze) anwendbar sind, und ordnet sie auf beiden Seiten einer imaginären Trennlinie an.

Entscheidend für die Einbettung der Elemente in das Modell sind die nach den TF\*IDF–Werte skalierten Merkmale der positiv und negativ bewerteten Datensätzen. Vor der Berechnung der TF\*IDF–Werte findet die Datenbereinigung statt, wobei sogenannte Stoppwörter und andere Wörter, die für den Vergleich irrelevant sind, von Datensätzen entfernt werden.

Aufgrund der üblicherweise begrenzten Anzahl positiv etikettierter Trainingsdaten wird eine mehrfache Kreuzvalidierungsmethode angewendet. Diese Methode teilt die gesamten Trainingsdaten in gleichmäßige Abschnitte auf, die als Validierung– und Trainingsdaten verwendet werden. Das trainierte Modell sollte nun in der Lage sein, die charakteristischen Merkmale (Wortformen) für positiv und negativ etikettierte Daten zu trennen und sie als eine sortierte Liste der entsprechenden Zeichengruppen darzustellen.

Das trainierte Modell wird durch die Vorhersagefunktion erweitert. [[25]](#footnote-27) Obwohl die Funktion in einigen Fällen scheitert, kann sie dennoch als zuverlässiges Werkzeug betrachtet werden. Der genaue Grund für das Scheitern der Vorhersagefunktion in einigen Fällen ist nicht eindeutig zu ermitteln. Es ist jedoch anzunehmen, dass dies teilweise an der Methode liegt, die Ähnlichkeit zwischen dem Modell und den Messdaten abzugleichen. Die Sigmoid–Funktion, die erfolgreich zur Berechnung der Distanz zwischen dem Modell und den Messdaten eingesetzt werden kann, könnte hierbei möglicherweise behindert werden, da im geprüften Textteil sowohl negative als auch positive Merkmale gleichzeitig vorhanden sein können.

Unter Berücksichtigung des oben genannten Faktors scheint es viel effektiver zu sein, die Daten ausschließlich anhand der mit den höchsten Koeffizienten versehenen positiven Merkmale zu untersuchen. Dadurch werden die Datensätze nicht mehr über das Modell verglichen, sondern über das aus dem Modell extrahierte prototypische Metamotiv bzw. Metaepisode. Dieses funktioniert wie eine Suchabfrage an die Dokumentensammlung.

Ein Bild, das Screenshot, Reihe, Rechteck, Diagramm enthält.

Automatisch generierte Beschreibung

Abb. 6

Das Extrahieren der Merkmale erfolgt durch die Berechnung des Schwellenwertes für die Topklassifikationsmerkmale. Dieser Wert ist an der Stelle der Merkmal–Kette zu suchen, wo die Gewichte der positiven und negativen Merkmale bzw. Koeffizienten abgeglichen werden.

Nach dem graphischen Verfahren kann man diese Stelle wie folgt lokalisieren: Die Abbildung 6 zeigt die Schwingung zwischen den maximalen und minimalen Koeffizienten der negativ und positiv eingeschätzten Klassifikationsmerkmale. Die für beiden Kategorien geltende maximale Zahl der Merkmale ›z‹ ist eine Variable und kann entsprechend der Situation angepasst werden, auf diesem Bild beträgt sie 200. Blaue Spalten stellen den positiven Merkmalen dar, die roten Spalten visualisieren hingegen die negativen Merkmale (Wegen der hohen Anzahl überlappen sich die Merkmale (siehe schwarze Wortwolke im unteren Bereich der Grafik). Für die Visualisierung der gesuchten Stelle wird die vorhandene Grafik an dem Treffpunkt von roten und blauen Spalten geteilt und die linke Hälfte so weit nach rechts verschoben, bis die kleinste rote Spalte unter der größten blauen Spalte platziert wird. Beachtet man die Größe von blauen und roten Spalten kann man leicht feststellen, an welcher Stelle sie übereinstimmen (s. Abb. 6.1).

Ohne graphische Darstellung kann man derselben Schwellenwert anhand folgender Formel berechnen:

Kommt die Gleichung zustande: , wobei , so ist der Schwellenwert gleich .

Da die Topklassifikationsmerkmale nichts anderes darstellen als eine einfache Liste von positiv bewerteten Elementen bzw. Wörtern, ist ihre Bildung mit Hilfe einer einfachen Stufenfunktion möglich (entspricht der grünen Grafik auf der Abb. 6.1):

Ein Bild, das Screenshot, Reihe, Rechteck, parallel enthält.

Automatisch generierte Beschreibung

Abb. 6.1

Die Liste der auf diese Weise festgelegten Merkmale wird als prototypisches Metamotiv in einem Datensatz gespeichert und bei den Messdaten angelegt.

Die Erkennung des gesuchten Motivs in den Messdaten erfolgt durch die Klassifikation auf Basis der Kosinus-Ähnlichkeit. [[26]](#footnote-28) Dabei wird jeder Datensatz mit einem synthetischen Datensatz verglichen, der von uns oben als das prototypische Metamotiv definiert wurde.

# Test des Assistenten

Für den Test der Vorhersagefunktion des Assistenten werden wir folgendes Experiment durchführen. [[27]](#footnote-29) Zunächst erstellen wir ein Textkorpus aus Daten, die mit hoher Sicherheit dem Genre Märchen zuzuordnen sind. Hierfür nutzen wir die im TextGrid–Repositorium veröffentlichten deutschsprachigen Märchensammlungen.

Wir wählen die Märchentexte von Jacob und Wilhelm Grimm (215), Heinrich Pröhle (150), Karl Bartsch (47) und Ludwig Bechstein (52+81) aus und fügen sie in einem XML-TEI-Corpus zusammen. [[28]](#footnote-30) Zusätzlich werden 12 weitere Texte aus der Sammlung ›Europäische Märchen‹ aufgenommen.[[29]](#footnote-31) Aufgrund rechtlicher Gründe sind diese Texte in einer abgeleiteten Fassung dargestellt. Insgesamt umfasst unser Korpus somit 557 Texte.

Nun wählen wir ein Beispiel eines Märchentyp ›ATU 300 – Der Drachentöter‹ aus der Kollektion von Karl Bartsch aus und annotieren den Text anhand des Markups für die inhaltliche Annotation des Märchens.[[30]](#footnote-32)

Der ausgewählte Text stellt ein komplexes Märchen dar, das aus vier Typteilen besteht. Nach den Anfangsepisoden aus den Typen ›ATU 567‹ (vom Essen des außergewöhnlichen Vogelherzes bis zur Flucht) und ›ATU 303‹ (von Großziehen von zwei ähnlich aussehenden Brüdern bis zur Trennung der Brüder und Austausch der Lebensindikatoren) beginnt die Episode des Typs ›ATU 300 – Der Drachentöter‹. Diese umfasst den Inhalt von der Erkundigung der menschlichen Opfergabe bis zur Heirat der geretteten Prinzessin. Der Text wird weiter als Typ ›ATU 303‹ fortgesetzt. Dabei wird berichtet, wie der jüngste Bruder von der Hexe versteinert, aber dann mit Hilfe seines Bruders wieder erlöst wird.

Für das Experiment werden wir in diesem Märchen nur den Textteil des Typs ›ATU 300‹ erschließen, das heißt mit Hilfe des Markups labeln. Somit wird das Modell nur anhand positiv gelabelter Datensätze trainiert. Im Rahmen dieses Verfahrens werden aus allen gelabelten Datensätzen die Tokens extrahiert und ein Vokabular erstellt.

Dabei wird das Vokabular von sogenannten Stoppwörtern und anderen für den Vergleich unwesentlichen Wörtern bereinigt. Letztere gelten als seltene Wörter, die nur vereinzelt in den Datensätzen vorkommen und daher keine Rolle beim Vergleich dieser Datensätze mit anderen spielen. Für die Bereinigung des Vokabulars von vereinzelt stehenden Wörtern ist der Parameter ›min\_df=n‹ zuständig. Angesichts der geringen Anzahl der Datensätze sowie der Kürze der Liste der in diesen Datensätzen anzutreffenden Tokens, wird der Wert des ›min\_df‹ Parameters bei unserem Experiment auf den minimalen Wert ›min\_df=2‹ gesetzt. Das bedeutet, dass jedes im Vokabular erfasste Wort wenigstens in zwei verschiedenen Datensätzen vorkommen muss.

Als nächstes werden die im Vokabular vorhandenen Wortformen nach ihren TF\*IDF–Werten skaliert. Die TF (Term Frequency) eines Tokens wird berechnet als:

›Häufigkeit des Tokens in einem Datensatz / Gesamtzahl der Tokens im Datensatz‹

Die IDF (Inverse Document Frequency) eines Tokens wird berechnet als:

›Loge (Gesamtzahl der Datensätze + 1 / Anzahl der Datensätze mit dem gesuchten Token + 1) + 1‹

Anschließend werden die TF\*IDF-Werte durch das L2-Normalisierungsverfahren umskaliert. Das somit berechnete Koeffizient bestimmt den Wert jedes einzelnen Tokens im Vokabular. Die Rangordnung der Tokens im Vokabular ermöglicht es, die Anzahl der Tokens im Vokabular effektiv zu korrigieren.

Für die Annäherung zwischen dem erstellten Vokabular und dem Modell einer konkreten Geschichte ist die Segmentierung der Trainingsdaten in natürliche Handlungsabschnitte von besonderer Bedeutung. Sie stellt sicher, dass Wörter, die in zwei aufeinanderfolgende Handlungsabschnitte (Offenkundigkeit der Krise – Auseinandersetzung mit der Krise) wiederholt vorkommen, mit getrennten Datensätzen dargestellt werden. Dadurch erhalten sie höhere TF\*IDF-Gewichte oder werden überhaupt im Vokabular erfasst.

Im nächsten Schritt wird aus den restlichen (nicht gelabelten) Datensätzen im Textkorpus ein Messdatenkorpus aggregiert. In diesen Messdatenkorpus wird auch der zuvor erstellte Modell-Datensatz eingefügt. Anschließend werden alle anderen Datensätze aus dem Messdatenkorpus mit dem Modell-Datensatz verglichen, um vorherzusagen, welche von ihnen diesem am nächsten stehen.

Das Verfahren umfasst folgende Schritte:

1. Erstellung und Bereinigung des Vokabulars: Das Vokabular wird von Stoppwörtern und anderen für den Vergleich unwesentlichen Wörtern bereinigt.

2. Berechnung der TF\*IDF-Werte: Für jedes Wort im Vokabular werden die TF\*IDF-Werte berechnet.

3. Erstellung dünnbesetzter Matrizen: Für jeden Datensatz wird eine dünnbesetzte Matrix erstellt, die die TF\*IDF-Werte der enthaltenen Tokens repräsentiert.

4. Vergleich der Matrizen: Jede dieser Matrizen wird mit der Matrix des Modell-Datensatzes verglichen.

Die Berechnung der Ähnlichkeit zwischen den Datensätzen beruht auf der Formel der Cosinus-Ähnlichkeit, die wie folgt definiert ist:

Dabei entspricht ›ai‹ einem, in einem TF\*IDF Wert skalierte Merkmal (Token) aus dem Modell-Datensatz, ein ›bi‹entspricht hingegen einem in einem TF\*IDF Wert skalierte Merkmal (Token) aus dem beliebigen Mess-Datensatz. Je näher der Wert der berechneten Cosinus-Ähnlichkeit bei Eins liegt, desto plausibler ist die vorhergesagte inhaltliche Ähnlichkeit zwischen ›a‹ und ›b‹ Datensätzen. Diese Methode ermöglicht es, die Ähnlichkeit der Datensätze zu quantifizieren und diejenigen zu identifizieren, die dem Modell-Datensatz am nächsten stehen.

Die Datensätze mit den höchsten Koeffizienten, also die wahrscheinlichsten Vorhersagen, stehen an der Spitze der Liste. Anhand des Nachschlagewerks von Hans-Jörg Uther, dem ›Deutschen Märchenkatalog‹, lässt sich mühelos überprüfen, ob der Algorithmus alle im Textkorpus enthaltenen Texte bzw. ihre Teile gefunden hat, die zur gesuchten Märchenkategorie gehören (gemeint ist der Typ ›ATU 300 – Der Drachentöter‹).

Die Recherche im Katalog zeigt, dass in der Tat alle entsprechenden Texte durch den Algorithmus sicher erfasst wurden. Das sind wie folgt:

|  |  |
| --- | --- |
| Text im Katalog[[31]](#footnote-33) | Text ID im XML-Textkorpus[[32]](#footnote-34) |
| Bartsch 1879f. I, 474ff. | @n=“dummy20“ |
| Bechstein/Uther 1997 I, Nr.49 | @n=“dummy122“ |
| Grimm/KHM (1857) Nr. 60 | @n=“dummy484“ |
| Pröhle 1853, Nr. 4 | @n=“dummy234“ |
| Pröhle 1853, Nr. 5[[33]](#footnote-35) | @n=“dummy256“ |

In diesem Experiment gezeigte Verfahren ist insbesondere bei der Recherche und Erstellung der Trainingsdaten einzusetzen. Anhand der somit gewonnenen Trainingsdaten ist es möglich, die Vorhersage nicht nur zu präzisieren, sondern auch auf kleinere Inhaltselemente zu richten. Um diese Eigenschaft zu demonstrieren, führen wir ein weiteres Experiment durch.[[34]](#footnote-36)

Wir trainieren das Modell anhand der zuvor gelabelten Datensätze, darunter die oben erwähnten Texte des Typs ›ATU 551 Das Wasser des Lebens‹.[[35]](#footnote-37) Die ID-Zeichen dieser Texte im Textkorpus beginnen mit dem Suffix ›zyx\_‹. Nach der Aggregation eines provisorischen Textkorpus aus diesen und den im ersten Experiment gelabelten Datensätzen (Datensätze mit dem ID-Suffix ›dummy‹) können wir alle in diesem Korpus vorhandenen Markupelemente in Gestalt einer Liste abrufen und betrachten.

Als nächstes wählen wir aus dieser Liste ein oder mehrere Markupelemente aus und erstellen eine entsprechende Abfrage. Angenommen, wir suchen nach dem Motiv der Offenkundigkeit der Krankheit des Vaters. Die entsprechende Abfrage besteht aus den Attributen ›a551‹ und ›KRANKHEIT:‹. Der Doppelpunkt nach dem Wort ›KRANKHEIT‹ stellt sicher, dass ein weiterer Wortlaut wie ›KRANKHEIT\_beheben‹ in der Suche ignoriert wird.

Das Ergebnis der anschließenden maschinellen Analyse bestätigt die hohe Präzision der eingesetzten Vorhersagefunktion. Der erste Treffer entspricht dem Textteil mit dem gesuchten Motiv. Außerdem ist der Unterschied zwischen den Koeffizienten des ersten und der folgenden Datensätze so groß, dass sie zweifelsohne zwei kontrastiv unterschiedliche Vorhersagen darstellen müssen. Tatsächlich, wie der oben erwähnte Katalog bestätigt, sind im Textkorpus keine anderen Texte mit gleichem Inhalt zu erwarten.[[36]](#footnote-38)

# Tools für die Auswertung der Daten

Durch die inhaltliche Annotation wird es möglich, eine große Menge von Texten einheitlich und mit präziser Fokussierung auf die konkreten Forschungsfragen auszuwerten und Ergebnisse nachvollziehbar und anschaulich darzustellen. Hier betrachten wir Anwendungsbeispiel mit dem Ziel die Normalform eines Typs in Westeuropäische Kulturraum in der visualisierten Form auszuwerten.

Im Vorfeld der Auswertung muss es vorbestimmt werden ›a‹ nach welchem Typ bzw. Typencluster wir suchen und ›b‹ im Welchem Teil der Daten die Suche durchgeführt werden soll.

Angenommen, interessieren wir uns für die Normalform des Märchentypus ›a551 – Wasser des Lebens‹ im europäischen Märchenrepertoire[[37]](#footnote-39) (Deutsch, Österreichisch, Französisch, Italienisch, Ungarisch, Griechisch, Russisch).

Nach der Eingabe von entsprechenden Suchparameter greift der Assistent auf die annotierten Märchentexten mit passenden Identifikatoren zu und stellt die Auswertungsergebnisse in der Form von Graphen und dabei in der einfachen Liste der Kompositionsbau beteiligten Motive dar.[[38]](#footnote-40)

Der Graph zeigt ein zusammengeführtes Bild der in den verschiedenen Texten gegenseitig geketteten Motiven mit gesuchtem ›a‹ Attributwert. An den Peripherien der Kette können auch einige Motiven auftreten, die andere ›a‹Attributwerte haben. Diese stellen ihrerseits die Peripherien der Motivkette dar, die den mit dem gesuchten Typ kombinierten Typen ausmachen.[[39]](#footnote-41) Nach der Eingabe der Werten von solchen ›a‹ Attribute ist zu erwarten, dass der Graph der Beziehungen von Motiven innerhalb eines Typs zum Graphen der Beziehungen der Motive innerhalb eines Typenclusters umgewandelt wird.

Die Auswertung der Motivbestand in einem Typ bzw. Typcluster, wird automatisch durch die Aggregation der Bibliothek der dabei involvierten Motive gefolgt. Diese kann, solange es keine entsprechende Aufforderung vorliegt, eine einfache Reihe von Motiven ohne Hinweise auf jegliche Verbindung darstellen. Nach der Eingabe eines konkreten Motivs wird die Liste wie folgt reorganisiert: das eingegebene Motiv bekommt die Nummer Null und wird an seine Stelle in der Reihe von gesamten Motiven platziert, die auf der Zeitachse ihm nach vorne und nach hinter liegende Motive bekommen entsprechende Nummerierung mit Minus- und Pluszeichen und werden den Platz oberhalb bzw. unterhalb des mit Null bezeichneten Motivs übernehmen. Wird die Kette der Motive von einem Typ durch eine oder mehrere Motiven mit einem anderen ›a‹ Attribut getrennt so werden sie ohne Hinweis auf ihre Position in der Motivkette erfasst. Somit erhalten wir eine Liste von Motiven, die mit dem zusammengefassten Inhalt des ausgewerteten Typs übereinstimmt.

Angesichts der Unübersichtlichkeit, die durch die Wechsel der handelnden Personen hervorgerufenen Variationen der Motive mit sich bringen, ergibt sich die Empfehlung, das zusammenfassende Bild der Komposition des Märchens auf einige wenige Elemente zu verdichten. In diesem Sinne scheint es effektiv zu sein, das letzte Attribut aus dem annotierten Markupelement zu entfernen.

Beispiel: Im Typ ›a551‹ (=Attribut ›a‹) kann der Auftrag ›F‹ (=Attribut ›b‹) zur Beschaffung des Lebenswassers (=Attribut ›c‹) verschiedenen Protagonisten (Attribut ›d‹) erteilt werden, sei es dem kranken Vater (›rHH‹) oder dem Rat gebenden alten Mann (›rST‹). Dies führt automatisch dazu, dass bei der Zusammenfassung im Prinzip derselben Handlung zwei Markupelemente erstellt werden:

›a551:F:Heilmittel\_beschaffen:rHH‹ und ›a551:F:Heilmittel\_beschaffen:rST‹.

Durch die Abschaffung des Attributs ›d‹ bleibt nur ein Markupelement übrig, das trotz des Informationsverlustes in der Lage ist, ein gemeinsames Bild der Handlungskette zu vermitteln. Entsprechender Code findet man in GitHub-Repositorium.[[40]](#footnote-42)

Letztendlich hängt der wissenschaftliche Wert und die Plausibilität einer solchen Auswertung vollständig vom Grad der Repräsentativität der erfassten Forschungsdaten ab. In dieser Publikation beschränken wir uns darauf, die Funktionsweise des Codes zu demonstrieren. Die mit dem Einsatz des Codes erzielten Ergebnisse anhand des kaukasisches Erzählgutes werden separat veröffentlicht.

# Ausblick

Die vorliegende Arbeit untersucht allgemeine Regeln für die Organisation und Kombination von inhaltlichen Elementen in allen Ebenen des Märchens: Motiv, Episode, Vollgeschichte und gibt die formalisierte Beschreibung der erkannten Ordnungen. Auf Basis des, dadurch erfassten konzeptuellen Models des Genres ›Zaubermärchen‹ entwerfe ich ein Machine\_Learning\_Modell, das uns ermöglicht die kosten und zeitintensivste Aufgaben des vergleichenden Märchens zu automatisieren.

Ein besonderer Mehrwert der erzielten Forschungsergebnisse liegt in der Bereitstellung eines künstlichen Assistenten zur Erkennung inhaltlich-struktureller Elemente innerhalb einer unüberschaubaren Menge von Textdaten. Hervorzuheben ist dabei die Zusammenführung der Leistungen von zwei unabhängigen Algorithmen in einem einzigen. Nur durch diesen Ansatz ist es möglich, die Vorhersagefunktion erfolgreich zu verwenden, selbst wenn die nach dem Modell getrennten positiven und negativen Klassifikationsmerkmale in einem untersuchten Textabschnitt zusammengehalten werden.

Ein wichtiger Bestandteil der im Zuge der vorliegenden Arbeit bereitgestellten Infrastruktur ist das Modul zur visualisierten Darstellung der ausgewerteten Normalformen von einzelnen Typen. Besonders effektiv scheint die Anwendung der Visualisierungsmethode durch Digraphen zu sein. Das daraus entwickelte Bild vermittelt eine klare und lückenlose Vorstellung über die Popularität einzelner Typen und die Häufigkeitsfälle ihrer gegenseitigen Kombinationen.

Die Ergebnisse dieser Studie werden gegenwärtig praktisch angewandt, um zur Erschließung der nordkaukasischen Folklore beizutragen. Das von der DFG geförderte Projekt verfolgt das Ziel, ein elektronisches Datenkorpus zu erstellen und durch die Annotation der inhaltlich-strukturellen Eigenschaften der in diesem Korpus gesammelten Texte diese für die international-vergleichende Märchenforschung bereitzustellen. Als besondere Herausforderungen gelten dabei:

1. sprachlichen Vielfältigkeit des Materials[[41]](#footnote-43),

2. Einschränkungen bei der Anwendung der Open-Date Prinzipien.[[42]](#footnote-44)

Nach nur anderthalb Jahren Anwendung der in diesem Aufsatz erfassten theoretischen und praktischen Forschungsergebnisse zeichnet sich ihre Nachhaltigkeit deutlich ab. Besonders bemerkenswert ist dabei die Fähigkeit, sich neuen Entwicklungsanforderungen anzupassen. Dies wird deutlich in der Umstellung von der CSV-Datenkollektion auf ein XML-Datenkorpus sowie im Übergang von grundlegenden Segmentierungsverfahren von Episoden auf die Motivebene.

Bei fortgesetzter Anwendung dieser Arbeitsmethode lässt sich das Erreichen eines besonderen Synergieeffekts gut prognostizieren. Wie bereits angedeutet, entwickelt sich die vergleichende Märchenforschung sukzessiv, nach dem Vorbild eines selbstregulierenden Systems. Dies bedeutet unter anderem, dass jedes neu erschlossene Repertoire die vergleichende Analyse eines anderen Repertoires begünstigt; beide tragen wiederum zur Erschließung des nächsten bei, und so weiter. Unter Berücksichtigung dieses Aspekts werden die Ergänzungen der einzelnen infrastrukturellen Elemente, wie beispielsweise Markups oder synthetische Trainingsdaten, als beiläufiger Beitrag bei jedem neuen Einsatz der Arbeitsmethode betrachtet.

1. Vgl. Aarne 1913, S. 65. [↑](#footnote-ref-1)
2. Vgl. Krohn 1926, S. 29. [↑](#footnote-ref-2)
3. (Vgl. Anderson 1934 – 1940, S. 515. [↑](#footnote-ref-3)
4. Propp 1975, S. 25. [↑](#footnote-ref-4)
5. Propp 1975, S. 95. [↑](#footnote-ref-5)
6. Vgl. Uther 2004, Part 1–3. [↑](#footnote-ref-6)
7. Vgl. Uther 2004, Part 1–3. [↑](#footnote-ref-7)
8. Vgl. Uther 2004, Part 3, S. 174–396. [↑](#footnote-ref-8)
9. Vgl. Elguja Dadunashvili: kf. XSD Datenset. GitHub. 2024 [[online](https://github.com/edadunashvili/VerMa/tree/main/kf/vmf_c1.xsd)] [↑](#footnote-ref-11)
10. Vgl. Elguja Dadunashvili: kf. XSD Datenset. GitHub. 2024 [[online](https://github.com/edadunashvili/VerMa/tree/main/kf/vmf_d.xsd)] [↑](#footnote-ref-12)
11. Grimm (Hg.)1996 [1857], Bd. 2, Nr. 97; Vgl. Blécourt 2014, Sp. 509–514. [↑](#footnote-ref-13)
12. Gemeint werden Tupels mit geklonten und abgeleiteten Wortlauten. [↑](#footnote-ref-14)
13. ATU–Klassifikationssystem gilt als Abkürzung für den erstmalig im Jahr 1910 von Antti Aarne herausgegebenen Verzeichnis der Märchentypen. Der Katalog wurde 1928 und 1961 von Stith Thompson und 2024 von Hans-Jörg Uther überarbeitet: Aarne 1910, Thompson 1928, Thompson 1961, Uther 2004; MIT–Klassifikationssystem gilt als Abkürzung für: Thompson 1955–1958. [↑](#footnote-ref-15)
14. Declerck et al. (o.D.), passim; D'Huy 2019, passim. [↑](#footnote-ref-16)
15. Als jüngster Versuch sind folgende Publikationen zu nennen: Finlayson 2012, passim; Finlayson 2016, passim; Lendvai et al. 2010a; Lendvai et al. 2010b. [↑](#footnote-ref-17)
16. Eklund et al. 2023: passim [↑](#footnote-ref-18)
17. Vgl. Ginschel 1967, S. 291 [↑](#footnote-ref-19)
18. Hahn 1864, S. 74-85; Vgl. Aarne 1913, S. 3, Fn. 3 [↑](#footnote-ref-20)
19. Aarne 1910, S. iii) [↑](#footnote-ref-21)
20. Thompson Thompson 1955–1958, [↑](#footnote-ref-22)
21. Uther 2004, S. 174 [↑](#footnote-ref-23)
22. Vgl. Elguja Dadunashvili: Textkorpus. XML–Textkorpus. GitHub. 2024 [[online](https://github.com/edadunashvili/VerMa/tree/main/Textkorpus.xml)] [↑](#footnote-ref-24)
23. Dafür werden Abkürzungen ›cc‹ für die freigegebene Daten und ›cr‹vfür die mit dem Copyright geschützte Texte. [↑](#footnote-ref-25)
24. Vgl. Elguja Dadunashvili: erthaos\_11. Python Skript. GitHub. 2024 [[online](https://github.com/edadunashvili/VerMa/tree/main/erthaos-11.ipynb)] [↑](#footnote-ref-26)
25. Vgl. Elguja Dadunashvili: erthaos-11. Python Skript. GitHub. 2024 [[online](https://github.com/edadunashvili/VerMa/tree/main/erthaos-11.ipynb)] (Zellen #131-141). [↑](#footnote-ref-27)
26. Vgl. Elguja Dadunashvili: erthaos-11. Python Skript. GitHub. 2024 [[online](https://github.com/edadunashvili/VerMa/tree/main/erthaos-11.ipynb)], (Zellen #142-153). Als vorbildlich gilt die Anwendung von Needham 2016. [↑](#footnote-ref-28)
27. Vgl. Elguja Dadunashvili: erthaos-11\_experiment\_1. Python Skript. GitHub. 2024 [[online](https://github.com/edadunashvili/VerMa/tree/main/erthaos-11_experiment_1.ipynb)] [↑](#footnote-ref-29)
28. TextGrid Repository [[online](https://textgridrep.org/)] [↑](#footnote-ref-30)
29. Uther 2005 [↑](#footnote-ref-31)
30. Bartsch 1879–1880 [[online](https://hdl.handle.net/11858/00-1734-0000-0005-D975-3)] [↑](#footnote-ref-32)
31. Vgl. Uther 2015, S. 71f. [↑](#footnote-ref-33)
32. Vgl. Elguja Dadunashvili: Textkorpus. XML–Textkorpus. GitHub. 2024 [[online](https://github.com/edadunashvili/VerMa/tree/main/Textkorpus.xml)] [↑](#footnote-ref-34)
33. Der Text stellt eine Kombination von Typen 303 und 300 dar. Im Katalog von Uther nur bei der Erschließung des Typs ATU 303 erfasst, vgl. Uther 2015, S. 74f. [↑](#footnote-ref-35)
34. Vgl. Elguja Dadunashvili: erthaos-11\_experiment\_2. Python Skript. GitHub. 2024 [[online](https://github.com/edadunashvili/VerMa/tree/main/erthaos-11_experiment_2.ipynb)] [↑](#footnote-ref-36)
35. Vgl. Uther 2005 [↑](#footnote-ref-37)
36. Vgl. Uther 2015, S. 71f. [↑](#footnote-ref-38)
37. Entspricht den in der DVD-ROM Ausgabe der ›Europäische Märchen‹ zusammengetragene Texte unter dem Typnummer ATU 551: Uther 2005. [↑](#footnote-ref-39)
38. GitHub: edadunashvili/VerMa/Auswertung\_abcd.ipynb [↑](#footnote-ref-40)
39. Stellt ein Mitglied aus den gesuchten Motiven den Anfang oder das Ende eines Textes dar, so wird dieses durch extra Zeichen „Anf-“ und „End-“ gekennzeichnet. [↑](#footnote-ref-41)
40. GitHub: edadunashvili/VerMa/Auswertung\_abc.ipynb [↑](#footnote-ref-42)
41. Auf dem heutigen Territorium der Föderationsrepublik Dagestan werden ungefähr 25 einheimische Sprachen gesprochen. [↑](#footnote-ref-43)
42. Als Ausweg aus dieser Situation bleibt die Umwandlung der Originaltexte in abgeleitete Forschungsdaten. Häufig verwenden wir dabei Texte, die von Stoppwörtern bereinigt wurden. Referenzen zu den Quelldaten sind bei der Freigabe solcher Ableitungen immer routinemäßig nachweisbar. [↑](#footnote-ref-44)